THE ANNALS
of
MATHEMATICAL STATISTICS

FOUNDED AND EDITED BY H. C. CARVER, 1930-1938
EDITED BY S. S. WILKS, 1938-1949

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

VOLUME XXI

1950

THE ANNALS OF MATHEMATICAL STATISTICS

EDITOR
T. W. ANDERSON

ASSOCIATE EDITORS

R. C. BOSE
W. FELLER
M. A. GIRSHICK
E. L. LEHMANN
ALEXANDER M. MOOD
JOHN W. TUKEY

WITH THE COOPERATION OF

M. S. BARTLETT
DAVID BLACKWELL
GEORGE W. BROWN
HARALD CRAMÉR
WILLIAM G. COCHRAN
J. F. DALY
W. EDWARDS DEMING
J. L. DOOB
PAUL S. DWYER
CHURCHILL EISENHART
T. E. HARRIS
PAUL G. HOEL
HAROLD HOTELLING
HOWARD LEVENE
WILLIAM G. MADOW
H. B. MANN
FREDERICK MOSTELLER
J. NEYMAN
H. E. ROBBINS
S. N. ROY
HENRY SCHEFFÉ
WALTER A. SHEWHART
A. WALD
JACOB WOLFOWITZ
MAX A. WOODBURY

Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland

INSTITUTE OF MATHEMATICAL STATISTICS

General Office: Business Administration Building, University of Michigan, Ann Arbor, Michigan
C. H Fischer, Secretary-Treasurer
This address should be used for all communications concerning membership, subscriptions, changes of address, back numbers, etc., but *not* for editorial correspondence. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue.

Editorial Office: Department of Mathematical Statistics, Columbia University, New York 27, New York
T. W. Anderson, Editor
Manuscripts should be submitted to this address; each manuscript should be typewritten, double-spaced with wide margins, and the original copy should be submitted (preferably with one additional copy). Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be professionally drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.
Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

Subscription Price: $10.00 per year inside the Western Hemisphere; $5.00 elsewhere. Single issues $3.00. Back numbers are available at $10.00 per volume or $3.00 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC., BALTIMORE, MARYLAND, U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879.

Copyright, 1950, by the Institute of Mathematical Statistics.

CONTENTS OF VOLUME 21

ARTICLES

MISCELLANEOUS

SOME PRINCIPLES OF THE THEORY OF TESTING HYPOTHESES[1]

By E. L. Lehmann

University of California, Berkeley

Introduction:

1. The likelihood ratio principle. The development of a theory of hypothesis testing (as contrasted with the consideration of particular cases), may be said to have begun with the 1928 paper of Neyman and Pearson [16]. For in this paper the fundamental fact is pointed out that in selecting a suitable test one must take into account not only the hypothesis but also the alternatives against which the hypothesis is to be tested, and on this basis the likelihood ratio principle is proposed as a generally applicable criterion. This principle has proved extremely successful; nearly all tests now in use for testing parametric hypotheses are likelihood ratio tests, (for an extension to the non-parametric case see [33]), and many of them have been shown to possess various optimum properties.

At least in the parametric case the likelihood ratio test has a number of desirable properties. Among these we mention:

(i) Frequently it is easy to apply and leads to a definite and reasonable test.

(ii) If the sample size is large, and if certain regularity conditions are satisfied an approximate solution can be given for the distribution problems that arise in the determination of size and power of the test (Wilks [32], Wald [25]). In fact, if the likelihood ratio is denoted by λ, $-2 \log \lambda$ approximately has a central χ^2-distribution under the hypothesis, a non-central χ^2-distribution under the alternatives. The number of degrees of freedom in these distributions equal the number of constraints imposed by the hypothesis.

(iii) As was shown by Wald [25], under certain restrictions the likelihood ratio test possesses various pleasant large sample properties.

In view of this, one may feel that the likelihood ratio principle, although perhaps not always leading to the optimum test, is completely satisfactory, and that a more systematic study of the problem of test selection is not necessary. Unfortunately, against the pleasant properties just mentioned there stands a very unpleasant one. Cases exist, in which the likelihood ratio test is not only unsatisfactory but worse than useless, and hence the likelihood ratio principle is not reliable. Examples of this kind were constructed independently by H. Rubin and C. Stein; the following is Stein's example.

[1] Parts of this paper were presented in an invited address at the meeting of the Institute of Mathematical Statistics on Dec. 30, 1948, in Cleveland, Ohio.

Let X be a random variable capable of taking on the values $0, \pm 1, \pm 2$ with probabilities as indicated:

	-2	2	-1	1	0
Hypothesis H:	$\frac{\alpha}{2}$	$\frac{\alpha}{2}$	$\frac{1}{2} - \alpha$	$\frac{1}{2} - \alpha$	α
Alternatives:	pC	$(1-p)C$	$\frac{1-C}{1-\alpha}\left(\frac{1}{2} - \alpha\right)$	$\frac{1-C}{1-\alpha}\left(\frac{1}{2} - \alpha\right)$	$\alpha \frac{1-C}{1-\alpha}$

Here α, C are constants, $0 < \alpha \leqq \frac{1}{2}$, $\frac{\alpha}{2-\alpha} < C < \alpha$, and p ranges over the interval [0, 1].

It is desired to test the hypothesis H at significance level α. The likelihood ratio test rejects when $X = \pm 2$, and hence its power is C against each alternative. Since $C < \alpha$, this test is literally worse than useless, for a test with power α can be obtained without observing X at all, simply by the use of a table of random numbers. It is worth noting that the test, which rejects H when $X = 0$, has power $\alpha \dfrac{1-C}{1-\alpha} > \alpha$, so that a reasonable test of the hypothesis in question does exist.

The existence of such examples gives added importance to the problem of developing a systematic theory of hypothesis testing. It is the purpose of the present paper to give a brief survey of the work done on some aspects of such a theory and to indicate certain extensions and modifications of the existing theory. Some examples and applications will be considered. These will be restricted to parametric problems. For applications to testing non-parametric hypotheses see [12].

The results of sections 5 and 8 were obtained jointly by Gilbert Hunt and Charles Stein in 1945. They have not been published and were communicated to me by Professor Stein. I should like to express to him my gratitude for acquainting me with this material and for giving me permission to include it in this paper. I should also like to acknowledge my indebtedness to Professor Henry Scheffé who read the manuscript and made many helpful suggestions.

2. Formulation of the problem. The problem of testing a statistical hypothesis was formulated by Neyman and Pearson [18] as follows.

A random variable X is known to be distributed over a space $\mathfrak{X}$ according to some member of a family of probability distributions $\{P_\theta^X\}$, $\theta \in \Omega$. It will be assumed here that there is specified an additive class $\mathfrak{B}$ of sets in $\mathfrak{X}$, and that the probability distributions P_θ^X are probability measures defined over $\mathfrak{B}$. All sets or real valued functions mentioned in this paper will be assumed measurable $\mathfrak{B}$ unless otherwise stated. If $B \in \mathfrak{B}$, we shall write for the measure assigned to B by P_θ^X interchangeably $P_\theta^X(X \in B)$, $P_\theta^X(B)$, and if there is no possibility of confusion, $P_\theta(B)$. Throughout most of the paper it will be assumed that the probability measures P_θ^X are absolutely continuous with respect to a

given sigma finite measure μ defined over $\mathfrak{B}$, so that there exist non-negative functions f_θ such that

$$P_\theta(B) = \int_B f_\theta(x)\, d\mu(x). \tag{2.1}$$

We shall then say that $f_\theta(x)$ is a generalized probability density *w.r.* to μ.

A statistical hypothesis H specifies a subset ω of Ω, and states that the distribution of X is some P_θ^X with $\theta \in \omega$. A test of H is any subset w of $\mathfrak{X}$, the convention being that H is rejected if the observed value x of X is in w, and that in the contrary case H is accepted. The selection of w is to be made as follows. A number α is given, $0 < \alpha < 1$, the level of significance, and w must be such that

$$P_\theta(w) = \alpha \text{ for all } \theta \in \omega. \tag{2.2}$$

Subject to this restriction it is desired to maximize $P_\theta(w)$ for θ in $\Omega - \omega$. The interpretation of these conditions is immediate. Since $P_\theta(w)$ is the probability of rejecting H computed under the assumption that P_θ^X is the distribution of X, equation (2.2) states that the probability of rejecting H is to be α (usually some small number such as .01 or .05) whenever H is true. Similarly the second condition expresses the fact that H is to be rejected with high probability when θ is in $\Omega - \omega$.

Naturally the second condition is not to be taken literally but rather as a loosely stated principle of choice. For in general there will exist a unique set w maximizing $P_{\theta_1}(w)$ for any given $\theta_1 \in \Omega - \omega$, but this w will change with θ_1. The condition has a clear meaning only in the case that the set $\Omega - \omega$ contains only a single point, and in a few special problems in which the same set w maximizes $P_\theta(w)$ for all $\theta \in \Omega - \omega$. In the general case there are available two main methods for making the condition precise. One may restrict consideration to some class of "nice" tests, so that within this class the maximization of $P_\theta(w)$ can be achieved uniformly for $\theta \in \Omega - \omega$. Alternatively, instead of asking that a local optimum property hold uniformly, one may look for a test whose power function possesses some optimum property in the large. Both of these approaches have an element of arbitrariness. In the first, the selection of a class of nice tests, in the second, the choice of an appropriate optimum property. Fortunately, in a number of important special cases, both methods, for various reasonable definitions, lead to the same test.

Before proceeding with this development, we shall modify the formulation of the problem slightly. First, as has been pointed out by many writers, it seems more natural to replace (2.2) by

$$P_\theta(w) \leqq \alpha \text{ for all } \theta \in \omega. \tag{2.3}$$

Secondly, we shall permit "randomized" tests (see [11, 29]), that is, instead of demanding that the statistician decide for each value of x whether to accept or to reject H, we shall allow the possibility that for certain x the decision be

reached by means of some chance device such as a table of random numbers. By a test of H we shall therefore mean a function ϕ from $\mathfrak{X}$ to the interval [0, 1], with the convention that when x is the observed value of X some chance experiment with two possible outcomes R, $\bar{R}$ will be performed such that $P(R) = \phi(x)$, and that H will be rejected when the outcome is R and will otherwise be accepted. The case of a non-randomized test w clearly is obtained as a special case by taking for ϕ the characteristic function of the set w.

For a test ϕ the probability of rejection is given by

$$\int_{\mathfrak{X}} \phi(x)\, d\, P_\theta^X(x) = E_\theta \phi(X) \tag{2.4}$$

where E_θ denotes expectation computed with respect to the probability distribution P_θ^X. We therefore obtain the following formulation of the problem: To determine a test function ϕ ($0 \leqq \phi(x) \leqq 1$) which maximizes $E_\theta\, \phi(X)$, the *power* of ϕ against the alternative θ, for θ in $\Omega - \omega$ subject to the condition

$$E_\theta \phi(X) \leqq \alpha \text{ for all } \theta \in \omega. \tag{2.5}$$

In this connection it is convenient to use the term *"level of significance"* for the preassigned number α, and to define the *size* of the test ϕ as

$$\sup_{\theta \in \omega} E_\theta \phi(X). \tag{2.6}$$

Except in the trivial case that there exists a test of size $< \alpha$ whose power is 1 against all alternatives, the size of any optimum test (in fact, of any admissible test) equals the level of significance.

3. Testing against a simple alternative. A complete solution of the problem formulated in the last section is available only in the case that ω and $\Omega - \omega$ each contains only a single point, that is, in the case that both the hypothesis and the alternative are simple. The solution is then given by the fundamental lemma of Neyman and Pearson [18], which we may state in the following slightly more complete form.

THEOREM 3.1. *Let*

$$P_\theta(A) = \int_A f_\theta(x)\, d\mu(x). \tag{3.1}$$

(a) *For testing the hypothesis H: $\theta = \theta_0$ against the alternative $\theta = \theta_1$ at level of significance α, there exists a number k and a test ϕ of size α such that*

$$\begin{aligned} \phi(x) &= 1 \quad \textit{when} \quad f_{\theta_1}(x) > k\, f_{\theta_0}(x), \\ \phi(x) &= 0 \quad \textit{when} \quad f_{\theta_1}(x) < k\, f_{\theta_0}(x). \end{aligned} \tag{3.2}$$

(b) *If $f_{\theta_0}(x)$ and $f_{\theta_1}(x)$ are $\neq 0$ for all x in $\mathfrak{X}$, then a test ϕ is most powerful for testing H against $\theta = \theta_1$ if and only if it satisfies (3.2) except possibly on a set of $\mu-$ measure* 0[2]. (Note that the number k of (3.2) is essentially unique).

[2] Throughout the paper we shall consider two tests as equal if they differ only on a set of μ-measure 0.

The second half of the theorem may be paraphrased by saying that under the conditions stated the most powerful test is uniquely determined by (3.2) except on the set on which

$$f_{\theta_1}(x) = k\, f_{\theta_0}(x). \tag{3.3}$$

On this set the value of ϕ may be assigned arbitrarily provided the resulting test has size α. If in particular the set on which (3.3) holds has measure 0, the most powerful test is unique.

It should be mentioned that (3.1) is no restriction since any two probability measures P_1, P_2 defined over a common additive class can be represented in this form with $\mu = P_1 + P_2$. If the assumption of (b) is not satisfied, the theorem is still true in essence but some trivial modifications are necessary.

No such complete solution is available for the problem of testing a composite hypothesis against a simple alternative. However, as was shown in [11], this problem may in many cases be reduced to the one just considered. Let the hypothesis state that θ is an element of ω, and consider the simple alternative $\theta = \theta_1$. Suppose that an additive class of sets has been defined on ω (in most of the applications ω is a subset of Euclidean space, and the additive class is formed by the Borel sets contained in ω). Then for any probability distribution λ over ω,

$$h_\lambda(x) = \int_\omega f_\theta(x)\, d\lambda(\theta) \tag{3.4}$$

is a probability density function with respect to μ.

Under certain conditions to be stated below, the most powerful test ϕ_λ for testing the simple hypothesis H_λ that X is distributed with probability density h_λ against the alternative f_{θ_1} is also most powerful for testing the original hypothesis H against the same alternative. This is essentially the Bayes approach developed by Wald for his general decision theory, and in fact, under the conditions which we shall state, λ is a least favorable distribution over ω in the following sense. Let β_λ be the power of ϕ_λ against f_{θ_1}, and for any distribution λ^* over ω denote by H_{λ^*}, ϕ_{λ^*}, β_{λ^*} the associated hypothesis, the most powerful test for testing it against f_{θ_1}, and the power of this test respectively. Then λ is said to be least favorable if for all λ^*

$$\beta_\lambda \leqq \beta_{\lambda^*}. \tag{3.5}$$

THEOREM 3.2. *Suppose there exists a probability distribution λ over ω such that the most powerful test ϕ_λ of size α for testing H_λ against f_{θ_1} is of size α also with respect to the original hypothesis H. Then*

(i) *ϕ_λ is most powerful for testing H against f_{θ_1};*

(ii) *λ is a least favorable distribution.*

Also, if ϕ_λ is the unique most powerful test for testing H_λ against f_{θ_1}, it is the unique most powerful test for testing H against f_{θ_1}.

These results are essentially contained in Wald's work (see for example theorem 4.8 of [26]).

There are many trivial applications of this theorem to finding most powerful tests of one-sided hypotheses concerning a single real-valued parameter, such as testing $H: p \leqq p_0$ against $p = p_1(p_0 < p_1)$ when X has a binomial distribution with parameter p. As is well known, it turns out in a number of these cases that the most powerful tests are in fact uniformly most powerful against the one-sided class of alternatives.

In [11] Theorem 3.2 was used to determine most powerful tests of certain hypotheses concerning normal distributions. As an example consider the case that $X_1, \cdots, X_n$ are independently normally distributed with common mean ξ and variance σ^2. Denote by H_1 and H_2 the hypotheses $\sigma = 1$ and $\xi = 0$ respectively, and let the alternative be: $\xi = \xi_1$, $\sigma^2 = \sigma_1^2$. Then the most powerful test of H_1 rejects if

$$(3.6) \qquad \begin{aligned} &\Sigma(x_i - \xi_1)^2 < k_1 \quad \text{when} \quad \sigma_1 < 1, \\ &\Sigma(x_i - \bar{x})^2 > c_1 \quad \text{when} \quad \sigma_1 > 1, \end{aligned}$$

and accepts otherwise. Here k_1 and c_1 depend only on the level of significance, that is, are independent of ξ_1, σ_1. If $\xi_1 > 0$, the most powerful test for testing H_2 rejects if

$$(3.7) \qquad \begin{aligned} &\Sigma(x_i - b)^2 \leqq k_2 b^2 \quad \text{when} \quad \alpha < \tfrac{1}{2}, \\ &\frac{\bar{x}}{\sqrt{\Sigma(x_i - \bar{x})^2}} \leqq c_2 \quad \text{when} \quad \alpha \geqq \tfrac{1}{2}, \end{aligned}$$

and accepts H_2 otherwise. Here k_2 and c_2 depend only on α, while b depends on ξ_1, σ_1 and α.

These results indicate that even when the class of alternatives is larger than in the above problems, some improvement over the standard tests may be possible provided good power is desired only against a narrow class of alternatives.

4. Sufficient statistics. Before treating the problem of composite alternatives, we shall consider an important simplification that can be obtained by making use of sufficient statistics. This notion was introduced by R. A. Fisher, and was further developed by J. Neyman [13] and in [2] and [10]. Consider any measurable partition of $\mathfrak{X}$. For any point x in $\mathfrak{X}$, let $t(x)$ be that set of the partition in which x lies. A set in the range of t is said to be measurable if the corresponding set of points x is an element of $\mathfrak{B}$. Denote the class of measurable t-sets by $\mathfrak{A}$. Then the *statistic* $T = t(X)$ is a random variable defined over $\mathfrak{A}$. Kolmogoroff has shown how for any $B \in \mathfrak{B}$ one can define the conditional probability $P(B \mid t)$ of B given $T = t$ uniquely up to a set of measure zero by the equation

$$(4.1) \qquad P(B \cap t^{-1}(A)) = \int_A P(B \mid t)\, dP^T(t) \quad \text{for all} \quad A \in \mathfrak{A}.$$

Suppose now that we are given a class $\mathfrak{F}$ of probability distributions for X, $\mathfrak{F} = \{P_\theta^X\}$, $\theta \in \Omega$. Denote by $P_\theta(B \mid t)$ the conditional probability of B given

$T = t$ computed for the distribution P_θ^T. The statistic T is said to be a *sufficient statistic* for $\mathfrak{F}$ (or for θ) if for every $B \in \mathfrak{B}$ there exists a determination of $P_\theta(B \mid t)$ that is independent of θ.

According to the above definition of statistic, $t(x)$ is an element of a measurable partition. However, one may consider instead any function t^* for which $t^*(x) = t^*(x')$ if and only if $t(x) = t(x')$, that is, any function that leads to this partition; the values that the function takes on are really immaterial. It will be convenient here to use this wider definition of statistic. For a rigorous treatment of some of the problems that will be referred to one needs to define an equivalence of statistics and to include in this definition the appropriate nullset considerations. A detailed account of these matters is given in [2] and [10].

From our present point of view tests are compared solely in terms of their power functions. On this basis two tests ϕ_1 and ϕ_2 may be considered equivalent if they have identical power, that is, if

$$E_\theta\phi_1(X) = E_\theta\phi_2(X) \text{ for all } \theta \in \Omega. \tag{4.2}$$

We can then state

THEOREM 4.1. *If T is a sufficient statistic for θ and $\phi(X)$ any test of a hypothesis concerning θ then there exists an equivalent test that is a function of T only.*

The proof of this theorem is immediate since

$$\psi(T) = E[\phi(X) \mid T] \tag{4.3}$$

is such a test.

It follows from Theorem 4.1 that we lose nothing by restricting consideration to tests based on a sufficient statistic.[3] The problem of determining whether or not some statistic is sufficient for a given family of distributions is simplified through the use of a criterion for sufficiency that can be checked on sight. This criterion is due to Neyman [13] who proved it in a somewhat special setting, and was recently proved in a very general form by Halmos and Savage [2]. It states that if $\mathfrak{F} = \{p_\theta\}$, $\theta \in \Omega$ is a family of generalized probability densities for X, then under certain mild restrictions a necessary and sufficient condition for $T = t(X)$ to be a sufficient statistic for $\mathfrak{F}$ is that $p_\theta(x)$ factors into one factor depending on θ but on x only through $t(x)$ and a second factor depending only on x.

The question arises as to which of various sufficient statistics to use. Since the purpose of introducing sufficient statistics is to reduce the complexity of a given statistical problem, one is led to seek a sufficient statistic that reduces the problem as far as possible and hence to the notion of a *minimal sufficient statistic*, a sufficient statistic T being *minimal* if it is a function of every other sufficient statistic (see [10]). It can be shown under fairly general conditions that a minimal sufficient statistic exists, and one can give an explicit construction for it.

---

[3] A justification for the use of sufficient statistics in the general statistical decision problem was given in [2].

As one would expect it turns out that the sufficient statistics commonly associated with various families of distributions are actually minimal. Thus for example, if $X_1, \cdots, X_n$ are independently normally distributed with common mean ξ and variance σ^2, the statistic $(\bar{X}, \Sigma(X_i - \bar{X})^2)$ is a minimal sufficient statistic for $\theta = (\xi, \sigma^2)$. If $X_1, \cdots, X_n$ are independently uniformly distributed over $(0, \theta)$, $\max(X_1, \cdots, X_n)$ is the minimal sufficient statistic for θ. If $\mathfrak{F}$ is the family of distributions according to which $X_1, \cdots, X_n$ are identically independently distributed according to an arbitrary univariate distribution (or according to an arbitrary probability density with respect to a fixed univariate measure), then the minimal sufficient statistic is obtained by defining for each point $x = (x_1, \cdots, x_n)$ the set $t(x)$ as the set of points obtainable from x by permutation of coordinates. Alternatively one can define it by $t(x_1, \cdots, x_n) = (\Sigma x_i, \Sigma x_i^2, \cdots, \Sigma x_i^n)$.

5. The principle of invariance. The notion of invariance was introduced into the statistical literature in the writings of R. A. Fisher, Hotelling, Pitman [20] and others, in connection with various special problems. A general formulation was given by Hunt and Stein who, in an unpublished paper [5], utilized this notion to find most stringent tests, and who obtained the examples of uniformly most powerful invariant tests that will be given below. The point of view in the present section is different from theirs however, since here invariance will only be considered as an intuitively appealing restriction that one may wish to impose on statistical tests.

We shall begin by considering an example. Suppose it were known that the height of people is distributed about a known mean, which for convenience we shall take to be zero, either according to a normal or to a Cauchy distribution, with unknown scale factor so that either

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}\,\theta} \exp\left(-\frac{x^2}{2\theta^2}\right) \qquad 0 < \theta < \infty \tag{5.1}$$

or

$$f_\theta(x) = \frac{\theta}{\pi} \frac{1}{\theta^2 + x^2}, \qquad 0 < \theta < \infty. \tag{5.2}$$

Suppose we wish to test from a sample $X_1, \cdots, X_n$ the hypothesis H that the true probability density belongs to the first of these classes against the alternative that it belongs to the second. Then it seems desirable that the decision of whether or not to accept H should be independent of the scale adopted for measuring the heights. For otherwise one worker expressing his data in feet might reject H while another worker using the same data but expressing them in inches would reach the contrary decision (In this connection see for example [34], p. 104) A "nice" test function ϕ therefore would be independent of the choice of scale, i.e., it would satisfy the condition

(5.3) $\phi(cx_1, \cdots, cx_n) = \phi(x_1, \cdots, x_n)$ for all $c > 0$ and for all $(x_1, \cdots, x_n)$ except possibly on a set N, independent of c and of measure zero.

On analyzing this problem one is led to the following observation. Multiplying each of the random variables $X_1, \cdots, X_n$ by the same constant leaves both ω and $\Omega - \omega$ invariant, i.e., if the X's are normally distributed with zero mean and arbitrary scale so are $cX_1, \cdots, cX_n$, and analogously for the Cauchy distributions. It is this fact that makes it so desirable to have ϕ invariant under multiplication of the x's by a common constant.

More generally consider measurable 1:1 transformations g of $\mathfrak{X}$ into itself, and let $Y = gX$. Suppose that when X is distributed according to $\theta \in \omega$, Y is distributed according to $\theta' \in \omega$—we shall then write $\theta' = \bar{g}\theta$—and that as θ ranges over ω so does θ'. Suppose that the analogous condition is satisfied for $\Omega - \omega$, so that the problem of testing ω against $\Omega - \omega$ is left invariant under g. Now whether one expresses the observations in terms of X or in terms of Y is essentially a matter of choice of coordinates. The principle of invariance asks that if such a change of coordinates leaves the problem invariant, then it should also leave the test invariant, i.e., if G is a group of measurable 1:1 transformations of $\mathfrak{X}$ such that

$$\bar{g}\omega = \omega \text{ and } \bar{g}(\Omega - \omega) = \Omega - \omega \text{ for all } g \in G, \tag{5.4}$$

then ϕ should satisfy the condition

$$\phi(gx) = \phi(x) \text{ for all } g \in G, \tag{5.5}$$

and for all x except on a set N independent of g and such that $\mu(N) = 0$. If this condition were not satisfied, two workers, using the same data but expressing them in different coordinate systems might arrive at contrary conclusions.

As an example consider the general linear univariate hypothesis. In canonical form $X_1, \cdots, X_r; X_{r+1}, \cdots, X_s; X_{s+1}, \cdots, X_n$ are independently normally distributed with common variance. The means of the first s variables are unknown, the means of the last n-s variables are known to be zero. The hypothesis states that the first r means are zero. Adding arbitrary constants to each of the variables of the middle group leaves ω and $\Omega - \omega$ invariant. So does any orthogonal transformation of the first r variables, and any orthogonal transformation of the last $n - s$ variables. Finally, the problem is also left invariant when all of the variables are multiplied by the same constant. It is easy to see that a function ϕ is invariant under these transformations if and only if it is a function of

$$\sum_{i=1}^{r} x_i^2 \Big/ \sum_{i=s+1}^{n} x_i^2.$$

But, as is well known and easy to show, among all tests based on this statistic there is a uniformly most powerful one, namely the test that rejects H when

$$\sum_{i=1}^{r} x_i^2 \Big/ \sum_{i=s+1}^{n} x_i^2$$

is too large. Therefore, among all tests satisfying the condition of invariance the standard test is uniformly most powerful.

To formulate a corresponding reduction procedure in general, we define a function h on $\mathfrak{X}$ to be maximal invariant (under G) if it is invariant and if $h(x') = h(x)$ implies the existence of $g \epsilon G$ such that $x' = gx$. Then a function φ on $\mathfrak{X}$ is invariant under G if and only if it depends on x only through $h(x)$, that is, if there exists a function ψ such that $\varphi(x) = \psi[h(x)]$. Hence a necessary and sufficient condition for a test to be invariant under G is that it be based on the statistic $Y = h(X)$. The principle of invariance therefore reduces the problem from X to $Y = h(X)$. To determine the resulting statistical reduction, that is, the simplification of the parameter space, one may consider the group $\bar{G}$ of transformations over Ω induced by G. If $v(\theta)$ is a maximal invariant function under $\bar{G}$, it is easily shown that the distribution of Y depends only on $v(\theta)$. Hence under the principle of invariance any two θ-values with common $v(\theta)$ (that is, such that each can be obtained from the other by a transformation of $\bar{G}$) are identified. If in particular $v(\theta)$ is constant over ω, the hypothesis H, when expressed for Y, becomes simple, and there may even exist a uniformly most powerful invariant test.

Besides for the example already mentioned this is the case for Hotelling's T^2-problem and for the hypothesis specifying the value of a multiple correlation coefficient. Another example is obtained when $X_1, \cdots, X_n$ are independently identically distributed, each with probability density $p_\theta(x)$ where under H_i $p_\theta(x) = f_i(x - \theta)$, $(i = 0, 1)$, and where it is desired to test H_0 against H_1. One may also in this example replace the location parameter by a scale parameter or have both parameters present.

It may be worth noting that the likelihood ratio test is invariant under any transformation leaving the statistical problem invariant. In the problems concerning normal distributions mentioned above, when there exists a uniformly most powerful invariant test, it coincides with the likelihood ratio test. That this is not so in general can be seen from Stein's example given in section 1. There the problem remains invariant under multiplication of X by -1, and there exists a uniformly most powerful invariant test. However, the likelihood ratio test is instead uniformly least powerful.

For certain applications it is more useful to consider a somewhat weaker definition of invariance. We shall say that a function φ is *almost invariant* under a group G of transformations if for each $g \epsilon G$, $\varphi(gx) = \varphi(x)$ for all x except on a set N_g such that $\mu(N_g) = 0$. This definition differs from the previous one in that the null set N_g is now permitted to depend on g. It was shown by Hunt and Stein that under certain conditions on G, which are satisfied for the problems mentioned above, any almost invariant test is invariant.

We have indicated how for certain hypotheses one can find a group of transformations leaving the problem invariant, such that among all tests invariant under this group there exists a uniformly most powerful one. The question may be raised whether this approach is consistent, or whether there may exist some other group of transformations also leaving the problem invariant but leading to a different test. Also in problems where among all invariant tests there does

not exist a uniformly most powerful one, the question arises whether one is using the totality of transformations leaving the problem invariant, or whether perhaps one can reduce the problem further. It therefore seems of interest to determine the totality of transformations leaving a given problem invariant. This was carried out for a few simple problems in [8].

We finally mention a connection between the notions of invariance and sufficiency. Consider any problem in which the variables $X_1, \cdots, X_n$ are independently identically distributed under all distributions of Ω. Such a problem clearly is left invariant under any permutation of the variables. Actually, these transformations leave not only ω and $\Omega - \omega$ invariant but each point of Ω individually. No essential reduction of the problem is obtained since the maximal invariant statistic is a sufficient statistic. It is easily seen that this will always be the case when the transformations leave Ω pointwise invariant, but that in this way one does not obtain all sufficient statistics. These can be obtained, however, by considering more general transformations, where each point x of $\mathfrak{X}$ is transformed into the points of $\mathfrak{X}$ according to a probability distribution P_x.

6. The principle of unbiasedness. As a second principle of reduction we shall consider the principle of unbiasedness proposed by Neyman and Pearson. A test is said to be unbiased [19] if

$$P_\theta \text{ (rejecting } H) \geq \alpha \text{ for all } \theta \in \Omega - \omega.$$

This seems a desirable property for a test to have since it assures that there do not exist θ_0 in ω and θ_1 in $\Omega - \omega$, for which

$$P_{\theta_0} \text{ (rejecting } H) > P_{\theta_1} \text{ (rejecting } H).$$

We shall therefore be concerned in this section with the totality of tests ϕ for which

$$\begin{aligned} E_\theta \phi(X) &\leq \alpha \quad \text{for all} \quad \theta \in \omega \\ E_\theta \phi(X) &\geq \alpha \quad \text{for all} \quad \theta \in \Omega - \omega. \end{aligned} \tag{6.1}$$

For a number of important special cases there exists, among all tests satisfying (6.1), one that is uniformly most powerful in $\Omega - \omega$ and uniformly least powerful in ω. (The latter property is of course very desirable since when H is true one wants to reject it as rarely as possible.) This follows immediately from well known results concerning best similar tests since for the problems in question Ω is a subset of a Euclidean space and for any test ϕ, $E_\theta\phi(X)$ is a continuous function of θ. If then Λ is the set of points that are boundary points both of ω and of $\Omega - \omega$, it follows from (6.1) that

$$E_\theta\phi(X) = \alpha \text{ for all } \theta \in \Lambda, \tag{6.2}$$

i.e., that ϕ is similar for θ in Λ. But if among all tests satisfying (6.2) there exists one that is uniformly most powerful in $\Omega - \omega$ and uniformly least power-

ful in ω, it automatically satisfies (6.1) as is seen by comparison with the test $\phi(X) \equiv \alpha$.

As an example suppose that $X_1, \cdots, X_n$ are independently normally distributed with common mean ξ and common variance σ^2. If the hypothesis is $H_1 : \sigma \leq 1$ and the alternatives are $\sigma > 1$, the set Λ becomes the line $\sigma = 1$. As was shown by Neyman and Pearson [18], among all tests satisfying (6.2) with this Λ, the test that rejects H_1 when $\Sigma(x_i - \bar{x})^2 \leq k$ (where k is an appropriately chosen constant) is uniformly most powerful for θ in $\Omega - \omega$, and uniformly least powerful for θ in ω.

If instead we consider testing the hypothesis $H_2 : \sigma = 1$ against the alternatives $\sigma \neq 1$, we find that $\Lambda = \omega$, and our problem reduces to that of finding the best test among all those that are similar in ω and unbiased. As is well known, it turns out that rejecting when $\Sigma(x_i - \bar{x})^2 \leq k_1$ and when $\Sigma(x_i - \bar{x})^2 \geq k_2$ (where $k_1 < k_2$ are two appropriately chosen constants) is uniformly most powerful among all similar unbiased tests.

A third hypothesis concerning σ that might be of interest is $H_3 : \sigma_1 \leq \sigma \leq \sigma_2$. Here Λ consists of the two lines $\sigma = \sigma_1$ and $\sigma = \sigma_2$ and it is easy to show that the test that is uniformly most powerful in $\Omega - \omega$ and uniformly least powerful in ω rejects H_3 if and only if $\Sigma(x_i - \bar{x})^2 \leq c_1$ or $\Sigma(x_i - \bar{x})^2 \geq c_2$ where again $c_1 < c_2$ are two appropriately selected constants.

The question arises as to the connection of the principles of invariance and unbiasedness. Clearly if there exists a unique test ϕ that is uniformly most powerful unbiased, this test is invariant under any group G leaving the problem invariant. If then in addition there exists a uniformly most powerful invariant (under G) test, this must coincide with ϕ. Thus, if both principles lead to a unique optimum solution, these solutions coincide.

We have seen that frequently optimum unbiased tests can be obtained through a study of tests that are similar over certain sets in the parameter space. The totality of similar tests was obtained for a number of important problems by Neyman and Pearson. In his 1937 paper on confidence intervals [15] Neyman gave a general method for constructing similar regions with the help of sufficient statistics. Let T be a sufficient statistic for $\theta \in \Lambda$. The condition for ϕ to be similar with respect to Λ and of size α, is that

(6.3) $$E_\theta\phi(X) = E_\theta E[\phi(X) \mid T] = \alpha \text{ for all } \theta \in \Lambda,$$

i.e., that

(6.4) $$E_\theta\{E[\phi(X) \mid T] - \alpha\} = 0 \text{ for all } \theta \in \Lambda.$$

Clearly any test ϕ for which

(6 5) $$E[\phi(X) \mid t] = \alpha \text{ for almost all } t$$

is similar. This is the construction given by Neyman, and we shall say that a test ϕ satisfying (6 5) has the Neyman structure with respect to T. The question whether this exhausts the totality of similar tests is easily reduced to an

analytic problem the solution of which is known in many special cases This method was first employed by P. L. Hsu [3] for some problems concerning normal distributions, and was extended to other cases in [7]. The present general formulation was given by H. Scheffé and the author in [9] and [10]. We shall say that a family of distributions $\{P_\theta^T\}$, $\theta \in \Lambda$, is boundedly complete if

(i) $f(t)$ is bounded,

(ii) $E_\theta f(T) = 0$ for all $\theta \in \Lambda$

imply $f(t) = 0$ except on a set N with $P_\theta(N) = 0$ for all $\theta \in \Lambda$. Then we can state

THEOREM 6.1. *A necessary and sufficient condition for the totality of tests similar for Λ to have Neyman structure with respect to a sufficient statistic T is that $\{P_\theta^T\}$, $\theta \in \Lambda$, be boundedly complete.*

7. Tests whose power increases with the distance from the hypothesis. Frequently, even among the unbiased tests, there does not exist a uniformly most powerful one. The general univariate linear hypothesis with more than one constraint is an example of this situation. The following extension of the idea of unbiasedness may then be used to reduce the class of tests still further. Unbiasedness distinguishes between values of θ as they belong to ω or $\Omega - \omega$ However, one may further classify the points of $\Omega - \omega$ according to their "distance" from ω, and then ask of a test φ that the further be θ from ω the larger be the power $\beta_\varphi(\theta)$.

One possible such ordering of the alternatives is that induced by the envelope power function. Here the envelope power at θ (Wald [24]) is defined by

$$\beta_\alpha^*(\theta) = \sup_{\varphi \in \mathfrak{F}(\alpha)} \beta_\varphi(\theta) \tag{7 1}$$

where $\mathfrak{F}(\alpha)$ is the class of all tests φ with $E_\theta\varphi(X) \leq \alpha$ for all $\theta \in \omega$. Of two points θ_1, θ_2 one may then say that θ_1 is closer to ω than θ_2, equally close or less close, as $\beta_\alpha^*(\theta_1)$ is less than, equal to or greater than $\beta_\alpha^*(\theta_2)$. The distance of θ from ω is thus measured by the ease with which one can detect that the hypothesis is false when θ is the true parameter value.

When θ lies in a Euclidean space and $\beta_\varphi(\theta)$ is a continuous function of θ for all θ; as is the case in most applications, the condition that the power increase with β_α^* will usually imply that $\beta_\varphi(\theta_1) = \beta_\varphi(\theta_2)$ whenever $\beta_\alpha^*(\theta_1) = \beta_\alpha^*(\theta_2)$. In the case of the general linear hypothesis considered in section 5, for example, one would obtain the condition that the power be a function only of $\sum_{i=1}^{r} \xi_i^2/\sigma^2$ where $\xi_i = E(X_i)$. As was shown by P. L. Hsu [3], the standard (likelihood ratio) test is uniformly most powerful among all tests satisfying this condition. Analogous remarks apply to Hotelling's T^2- problem, and to the hypothesis specifying the value of the multiple correlation coefficient. The corresponding optimum properties in these cases were proved by Simaika [21].

It is interesting to compare the above condition with that of invariance.

This comparison yields nothing of interest if the totality of tests is considered. We may, however, restrict our attention to tests depending only on a sufficient statistic T. We already know that $\varphi(X)$ and $E[\varphi(X) \mid T]$ have identical power. In order to validate the comparison we wish to make, we state the following

LEMMA. *Let T be a sufficient statistic for $\theta \in \Omega$, and let G be a group of 1:1 transformations g on X leaving Ω invariant. Then if $\varphi(x)$ is invariant under G, $E[\varphi(X) \mid t]$ is almost invariant under G.*

We can now state the desired comparison in the following

THEOREM 7.1. *Let G be a group of 1:1 transformations on X, let $\bar{G}$ be the induced group of transformations on Ω, let $v(\theta)$ be maximal invariant under $\bar{G}$, and suppose that $\bar{G}$ leaves ω and $\Omega - \omega$ invariant. Suppose further that T is a sufficient statistic for Ω, and that $\{P_\theta^T\}$, $\theta \in \Omega$, is boundedly complete. Then a necessary and sufficient condition that the power of a test $\psi(T)$ be a function only of $v(\theta)$, is that $\psi(t)$ be almost invariant under G.*

This theorem is an immediate extension of some results of Wolfowitz [35].

Theorem 7.1 together with the results of section 5 proves that the standard tests of the general linear hypothesis, Hotelling's T^2-problem and the hypothesis concerning the multiple correlation coefficient possess the optimum property that was obtained for these problems by Hsu and Simaika, respectively. The method of proof indicated here is due to Wolfowitz [35].

8. Most stringent tests. We shall now turn to the third aspect of the theory: Optimum properties defined with reference to the whole class of alternatives, and attainable with no restrictions imposed on the class of tests. In the present section we shall consider the property of stringency. Wald [25] defines a test φ to be most stringent if it minimizes

$$\sup_{\theta \in \Omega - \omega} [\beta_\alpha^*(\theta) - \beta_\varphi(\theta)], \tag{8.1}$$

where β_α^* again denotes the envelope power, and β_φ the power of φ. The rationale of this definition is clear. The difference $\beta_\alpha^*(\theta) - \beta_\varphi(\theta)$ measures the amount by which the test falls short at the alternative θ of the power that could be attained against this particular alternative. A test φ is therefore most stringent if it minimizes its maximum shortcoming.

A theory of most stringent tests was developed by Hunt and Stein [5], who based it on the notion of invariance. Consider, as in section 5, a group G of measurable 1:1 transformations on $\mathfrak{X}$ leaving the problem invariant. Hunt and Stein obtained their results in connection with the following groups of transformations.

(i) $gx = x + c$, $-\infty < c < \infty$, x a real variable;
(ii) $gx = ax$, $0 < a$, x a real variable;
(iii) $gx = ax + c$, $0 < a$, $-\infty < c < \infty$, x a real variable;
(iv) the group of orthogonal transformations on a Euclidean space;
(v) any finite group

THEOREM 8.1. (Hunt and Stein). *If G is the direct product of a finite number of groups of types* (i)–(v), *and if G leaves the problem invariant, that is, if G satisfies* (5.4), *then there exists a most stringent test invariant under G.*

Actually, it is not necessary here to require that G be a direct product. The result holds also if the factoring of G is according to normal subgroups, where the normal subgroup at each stage and the final factor group are of the types mentioned. In the light of this one may omit type (iii) from the list since it has a normal subgroup of type (i) with factor group of type (ii).

The proof of Theorem 8.1 is based on the following lemma, which has applications to many related problems.

LEMMA (Hunt and Stein). *If G is a direct product of a finite number of groups of types* (i)–(v) *then given any function f over $\mathfrak{X}$ $(0 \leq f(x) \leq 1)$ there exists a function F $(0 \leq F(x) \leq 1)$ such that F is invariant under G, and*

$$\inf_{g \in G} \int f(gx)\varphi(x)\, d\mu(x) \leq \int F(x)\varphi(x)\, d\mu(x) \leq \sup_{g \in G} \int f(gx)\varphi(x)\, d\mu(x) \tag{8.2}$$

for all φ that are integrable μ.

It follows from Theorem 8.1 that if there exists a uniformly most powerful invariant test, this test is most stringent. In this way Hunt and Stein show, for example, (see in this connection section 5), that the likelihood ratio test of the general univariate linear hypothesis is most stringent. A question that is left open is the uniqueness of such a most stringent test.

In general, the possibility therefore remains that there might exist another most stringent test uniformly more powerful than the invariant one. In certain particular cases this possibility can be ruled out by the following consideration. Suppose that Ω is a subset of a Euclidean space and that every point of ω is a limit point of $\Omega - \omega$. Suppose further that for any test ϕ, $E_\theta\phi(X)$ is continuous in θ. Then clearly, if ϕ_1 is similar of size α for testing ω and ϕ_2 is of size $\leqq \alpha$ but not similar, ϕ_2 can not be uniformly as powerful as ϕ_1. Hence any test that is admissible among all similar tests of size α is also admissible among the totality of tests of size $\leqq \alpha$. Now admissibility among all similar tests is sometimes not too difficult to prove. For the likelihood ratio test of the general linear univariate hypothesis, for example, it is an immediate consequence of the properties of this test proved by Wald [23] and Hsu [4].

The following alternative method for obtaining most stringent tests is also mentioned by Hunt and Stein.

THEOREM 8.2. (Hunt and Stein). *Let $\Omega - \omega$ be partitioned into disjoint subsets Ω_δ such that $\beta^*_\alpha(\theta)$ is constant on each Ω_δ, and let φ_δ be the test that maximizes $\inf_{\theta \in \Omega_\delta} \beta_{\varphi\delta}(\theta)$. Then if $\varphi_\delta = \varphi$ is independent of δ, φ is most stringent.*

This result may be supplemented by the following method for finding tests that maximize $\inf_{\theta \in \omega_1} \beta_\varphi(\theta)$ over a given set of alternatives ω_1 (not necessarily satisfying the conditions imposed above on the Ω_δ's).

THEOREM 8.3. *Suppose additive classes of sets have been defined over* ω *and* ω_1, *and consider probability measures* λ *and* λ_1 *over* ω *and* ω_1. *Let the functions* $f_\theta(x)$ *be generalized probability densities with respect to* μ, *so that* $h(x) = \int_\omega f_\theta(x)\, d\lambda(\theta)$ *and* $h_1(x) = \int_{\omega_1} f_\theta(x)\, d\lambda_1(\theta)$ *are again probability densities with respect to* μ. *Let* φ *be the most powerful test of size* α *for testing the simple hypothesis* $H: h$ *against the simple alternative* h_1, *and suppose that the power of* φ *against* h_1 *is* β. *Then if*

$$E_\theta\varphi(x) \leq \alpha \quad \text{for all} \quad \theta \in \omega, \qquad E_\theta\varphi(x) \geq \beta \quad \text{for all} \quad \theta \in \omega_1, \tag{8.3}$$

φ *maximizes* $\inf_{\theta \in \omega_1} \beta_\varphi(\theta)$ *at level of significance* α.

This method, when applicable, has the advantage of giving the totality of most stringent tests (see in this connection Theorem 3.1) and hence of settling the question of admissibility. However, in many applications probability measures λ, λ_1 with the desired properties do not exist but instead only sequences $\lambda^{(n)}$, $\lambda_1^{(n)}$, which satisfy the conditions in the limit. In this case again only the weak conclusion is possible: The test obtained is most stringent but has not been proved admissible. (For an example in which the analogous method has been carried through in detail for an estimation problem, see [22]).

Actually, the two methods are closely related, as can be seen from the proof of the main lemma. In those cases in which there exists a group G giving the maximum possible reduction, the group $\bar{G}$ induces a partition of Ω (through the equivalence: $\theta_1 \sim \theta_2$ if there exists $\bar{g}$ such that $\theta_2 = \bar{g}\theta_1$), just into ω and the sets Ω_δ. (This is so mainly because, as was shown by Hunt and Stein, the envelope power remains invariant under any transformations that leave the problem invariant.) Then the measures λ, λ_δ over ω, Ω_δ respectively, which figure in the application of Theorems 8.2 and 8.3, become invariant measures over $\bar{G}$ through the obvious 1:1 mapping from ω and the Ω_δ's respectively to $\bar{G}$. Thus the second method will allow the strong conclusion when the group $\bar{G}$ involved in the first method possesses a finite invariant measure [types (iv) and (v)] but not if any of its factors are of type (i)–(iii).

To conclude this section we shall give an example where the method of invariance leads only to a partial reduction but where the solution may be completed by certain additional considerations. Suppose that $(X_1, \cdots, X_n)$ is a sample from a normal distribution with mean ξ and variance σ^2, both unknown, and that we wish to find the most stringent test of the hypothesis H: $\sigma = 1$ against the alternatives $\sigma \neq 1$. Theorem 8.1 reduces the problem to the statistic $Y = \Sigma(X_i - \bar{X})^2$, but among the tests of H based on this statistic there does not exist a uniformly most powerful one. It may also be shown [8] that no further reduction is possible by means of the method of invariance.

However, one may now consider the problem of finding the most stringent test based on Y. (The envelope power function $\beta^*(\xi, \sigma)$ that must be used

naturally is not the one for Y but that for the original problem.) From an argument given in [6] it follows that this test is of the form

$$\varphi_{k_1,k_2}\text{: reject when } Y < k_1 \text{ or } > k_2 ,$$

where k_1, k_2 are determined by the two conditions

(i) $P(\text{rejection} \mid \sigma = 1) = \alpha$,

(ii) $\sup_{\sigma<1} [\beta^*_\alpha(\xi, \sigma) - \beta_{\varphi_{k_1,k_2}}(\sigma)] = \sup_{\sigma>1} [\beta^*_\alpha(\xi, \sigma) - \beta_{\varphi_{k_1,k_2}}(\sigma)]$.

Here $\beta^*_\alpha(\xi, \sigma)$ is independent of ξ and can be obtained from a table of the χ^2-distribution (with n degrees of freedom for $\sigma < 1$ and n-1 degrees of freedom for $\sigma > 1$ as can be seen from (3.6)). Hence k_1 and k_2 can be computed fairly easily.

Another problem that may be treated in this way is the hypothesis of equality of variances for two normal samples. If the two samples are of equal size, there exists a uniformly most powerful invariant test for a suitable group of transformations. However, if the sample sizes are different the method of invariance reduces the problem only to $\Sigma(X_i - \bar{X})^2 / \Sigma(Y_i - \bar{Y})^2$, and the cut off points giving the most stringent test may be determined by an argument analogous to that given above.

This method may be extended to allow determination of most stringent test of hypotheses such as H: $\sigma_1 \leqq \sigma \leqq \sigma_2$ This requires a certain modification of Theorem 1 of [6], which is easily obtained. One finds agains that one may restrict consideration to a one-parameter family of tests (determined by a somewhat different condition than above), and that among these the most stringent test is obtained by the analogue of condition (ii) above.

If should be mentioned that the results of [6] apply also to the hypothesis specifying the value of the parameter in a binomial or Poisson distribution. This is easily seen since in either case the distributions of Ω are absolutely continuous with respect to a common sigma finite measure and since for the appropriate choice of this measure the generalised density is of the form assumed for the density in [6]. Hence in both the binomial and the Poisson case the most stringent test is determined by conditions analogous to (i) and (ii) above.

9. Tests that minimize the maximum loss. In the Neyman-Pearson theory one classifies the errors into two kinds: Rejecting the hypothesis when it is true, accepting it when it is false. One may however analyze the situation further and distinguish, say, between accepting when one or some other alternative is true. Thus one is led to introduce the losses that result in a given situation from the various possible errors, and to look for a test that, in an appropriate sense, minimizes the expected loss. This possibility was mentioned by Neyman and Pearson [17], and was taken as the starting point of his general theory by Wald (see for example [24]).

In order to stay within the framework of this exposition we shall here introduce losses only for the errors of accepting the hypothesis when it is false,

while still demanding that the probability of rejection when the hypothesis is true should not exceed α. Actually, there are many cases where this seems to be a reasonable formulation. For it frequently happens that the two types of error entail consequences of such completely different nature that the resulting losses cannot be measured on a common scale while usually the different errors of the same type are comparable.

We shall therefore assume that for each $\theta \in \Omega - \omega$ there is defined a $W(\theta)$, which measures the loss resulting from acceptance of H when θ is true. The risk which one runs by using a test φ, when $\theta \in \Omega - \omega$ is the true parameter value is given by the expected loss $R_\varphi(\theta) = W(\theta)\, E_\theta[1 - \varphi(X)]$. When a uniformly most powerful test exists for the hypothesis in question, this test also minimizes the expected loss uniformly for θ in $\Omega - \omega$. In the contrary case one may again restrict the class of tests in some way, so that within the restricted class there exists a uniformly most powerful test, and hence a test that uniformly minimizes the expected loss. Alternatively we may again consider some optimum property of the risk function $R_\varphi(\theta)$ as a whole. We shall here consider the minimax principle introduced by Wald, and seek a test, which, subject to $E_\theta\varphi(X) \leq \alpha$ for all $\theta \in \omega$, minimizes

$$\sup_{\theta \in \Omega - \omega} W(\theta)\, E_\theta[1 - \varphi(X)],$$

the maximum risk.

If one introduces losses also for the other type of error it is easy to see that for a suitably chosen loss function the definition of minimax expected loss coincides with that of stringency. It is therefore not surprising that the methods of the previous section can be extended to cover the problems considered in the present one. (They are actually much more general, and may be applied also, for example, to the problem of point estimation, and in fact to the general decision problem).

From the lemma of Hunt and Stein stated in the previous section we immediately obtain the following extension of Theorem 8.1.

THEOREM 9.1. *If G is a group of transformations leaving the hypothesis and the class of alternatives invariant, if G can be factored by normal subgroups into factors of types* (i)–(v), *and if the loss function $W(\theta)$ is invariant under G, then there exists a test φ invariant under G and minimizing*

$$\sup_{\theta \in \Omega - \omega} W(\theta)\, E_\theta[1 - \varphi(X)]. \tag{9.1}$$

It follows that when a uniformly most powerful invariant test exists, this test has the property of minimizing the maximum expected loss with respect to any invariant loss function. Thus Student's test, for example, minimizes the maximum risk for any loss function that depends only on $|\xi|/\sigma$.

Clearly the second method mentioned in section 8 can be extended in an analogous manner if in Theorem 8.2 one replaces the sets Ω_i by sets over which $W(\theta)$ is constant.

Again it may happen that the method of invariance does not reduce the problem sufficiently far but that the solution may be completed by other considerations. Let us once more consider the hypothesis $H: \sigma = 1$ of the previous section, and let us suppose that the loss function has the necessary invariance property, so that it is a function only of σ but not of the unknown mean. It follows from Theorem 9.1 that there exists a test minimizing the maximum risk, which is a function only of $Y = \Sigma(X_i - \bar{X})^2$. From [6] it is easily seen that a test φ_{k_1,k_2} which rejects when $Y < k_1$ or $> k_2$, has the desired property if its size is α and if in addition

$$(9.2) \qquad \sup_{\sigma<1} W(\sigma)E_\sigma[1 - \varphi(Y)] = \sup_{\sigma>1} W(\sigma)E_\sigma[1 - \varphi(Y)].$$

It follows that depending on the choice of $W(\sigma)$ the solution may be any member of the one-parameter family of tests φ_{k_1,k_2} of size α.

Under the conditions of Theorem 9.1, when a uniformly most powerful invariant test exists, this also maximizes the average power for a large class of weight functions. If there exists a common finite invariant measure over the sets Ω_δ in the sense indicated in section 8, the uniformly most powerful invariant test will maximize the average power with this measure as weight function, over Ω_δ for all δ. It follows that it maximizes the average power over $\Omega - \omega$ with respect to any weight function for which the conditional distribution over each Ω_δ is the above invariant measure. If the invariant measure over the Ω_δ's is not finite one can obtain analogous results with respect to a sequence of weight functions invariant in the limit. The results indicated here are much weaker than those obtained for the general linear univariate hypothesis by Wald [23] and Hsu [4] under the restriction to similar regions. However their results are no longer valid when this restriction is omitted.

10. Applications to sequential analysis. So far we have restricted consideration to the case that the hypothesis is to be tested on the basis of a preassigned experiment. However, frequently there is available for this purpose a large class of experiments, and the selection of an optimum experiment out of this class is part of the problem. We shall consider here only the following situation, which has recently been studied extensively (see Wald [28, 29]). There is given a sequence of random variables $X_1, X_2, \cdots$ whose joint distribution is known to belong to some family $\mathfrak{F} = \{P_\theta\}$, $\theta \in \Omega$; the hypothesis specifies some subfamily: $\theta \in \omega$. The X's are observed one by one, and the decision, whether or not to continue experimentation at any given stage, is allowed to depend on the observations taken up to that point. Thus the number n of observations that will be taken is a random variable whose distribution depends on θ. Usually, by an appropriate choice of stopping rule, there may be effected a considerable saving in the expectation of the number of observations necessary to achieve a given discrimination between hypothesis and alternatives. The problem is to determine the stopping rule and test that minimizes this expectation.

As we have seen in the previous sections the principal methods for obtaining

optimum tests consist in reducing the problem to that of testing a simple hypothesis against a simple alternative This basic problem was solved in the non-sequential case by Neyman and Pearson (Theorem 3.1). The solution of the much more difficult corresponding sequential problem was obtained for a large class of cases by Wald and Wolfowitz [31] in the following

THEOREM 10 1. *Let $X_1, X_2, \cdots$ be identically and independently distributed. It is desired to test the hypothesis that the common probability density of the X's is $f(x)$ against the alternative that it is $g(x)$. Given two numbers $0 < \alpha < \beta < 1$, there exists a test which, subject to the condition*

$$P\text{ (rejection} \mid f) \leq \alpha \tag{10.1}$$
$$P\text{ (rejection} \mid g) \geq \beta,$$

minimizes simultaneously $E_f(n)$ and $E_g(n)$, the expected number of observations computed for the distributions f and g. This test is given in terms of two numbers A and B by the following rule. After m observations have been taken,

$$\textit{reject if } \frac{g(x_1) \cdots g(x_m)}{f(x_1) \cdots f(x_m)} > A,$$

$$\textit{accept if } \frac{g(x_1) \cdots g(x_m)}{f(x_1) \cdots f(x_m)} < B,$$

$$\textit{take another observation if } B < \frac{g(x_1) \cdots g(x_m)}{f(x_1) \cdots f(x_m)} < A.$$

Here A and B are determined so that condition (10.1) *holds with the inequality signs replaced by equality.*

So as to be able to treat the various problems considered non-sequentially in the previous sections one would have to extend this theorem at least to the case that the variables $X_1, X_2, \cdots$ form a set of equivalent variables in the sense of de Finetti [1]. Instead, we shall here restrict ourselves to a few problems that can be solved on the basis of Theorem 10.1. All of the tests discussed below were derived from various points of view and some of their properties were discussed by Girshick in his important "Contributions to the theory of sequential analysis", *Annals of Math. Stat.*, vol. 17 (1946) pp. 123–143 and 282–298, and by Wald in his basic book on the subject [28].

It is convenient here to modify slightly the formulation of the problem of hypothesis testing. Let the parameter space Ω be divided into three sets, the set ω_0 specified by the hypothesis, the class of alternatives ω_1, and a region of indifference $\Omega - \omega_0 - \omega_1$ where we do not much care whether the hypothesis is accepted or rejected (see [28]). Let us denote the sequential random variable $(X_1, \cdots, X_n)$ by X. Then we wish to determine a sequential test φ, which, subject to

$$E_\theta\varphi(X) \leq \alpha \text{ for } \theta \in \omega_0 \tag{10.2}$$
$$E_\theta\varphi(X) \geq \beta \text{ for } \theta \in \omega_1,$$

minimizes $\sup_{\theta\epsilon\omega_0+\omega_1} E_\theta(n)$. (Actually, this is a rather artificial formulation. The natural requirement is the minimization of $\sup_{\theta\epsilon\Omega} E_\theta(n)$ but this is a much more difficult problem.) The reduction to the problem of testing a simple hypothesis against a simple alternative is achieved by the following obvious extension of Theorem 8.3.

THEOREM 10.2. *Let λ_0, λ_1 be distributions over ω_0, ω_1 respectively, and let φ be a test, which subject to*

$$\int_{\omega_0} E_\theta\varphi(X)d\lambda_0(\theta) \leq \alpha \tag{10.3}$$
$$\int_{\omega_1} E_\theta\varphi(X)d\lambda_1(\theta) \geq \beta,$$

minimizes $\sup_{i\epsilon\{0,1\}} \int E_\theta(n)\, d\lambda_i(\theta)$. *Then if φ satisfies* (10.2) *and*

$$E_\theta(n) \leq \sup_{i\epsilon\{0,1\}} \int E_\theta(n)\, d\lambda_i(\theta) \text{ for all } \theta \epsilon \omega_0 + \omega_1, \tag{10.4}$$

φ minimizes $\sup_{\omega_1+\omega_2} E_\theta(n)$ *subject to* (10.2).

As in section 3 we can make certain trivial applications to problems concerning a single real parameter such as testing the hypothesis $H: p \leq p_0$ against the alternatives $p \geq p_1$ $(p_0 < p_1)$, where p is the probability of success in a binomial sequence of trials. In this example condition (10.2) of Theorem 10.2 obviously is satisfied when λ_0 and λ_1 assign probability 1 to p_0 and p_1 respectively. Hence the probability ratio test for testing $p = p_0$ against $p = p_1$ has the desired properties, whenever (10.4) holds, that is, whenever $E_p(n)$ attains its maximum between p_0 and p_1.

The following is another example that may be solved in this manner. Let $X_1, X_2, \cdots; Y_1, Y_2, \cdots$ be independently normally distributed, all with unit variance and means $E(X_i) = \xi$, $E(Y_i) = \eta$. In order to test the hypothesis $H: \xi \geq \eta$ against the alternatives $\eta - \xi \geq \delta$ where $\delta > 0$ is given, a pair (X_1, Y_1) is observed. If after this observation experimentation continues another pair (X_2, Y_2) is observed, etc. In this case we may take for λ_0, λ_1 the distributions that assign probability 1 to the parameter points $(\xi, \eta) = (0, 0)$ and $\left(-\frac{\delta}{2}, \frac{\delta}{2}\right)$ respectively. Then the probability ratio after m observations is given by

$$\frac{\exp\left[-\frac{1}{2}\sum_{i=1}^{m}\left(x_i + \frac{\delta}{2}\right)^2 - \frac{1}{2}\sum_{i=1}^{m}\left(y_i - \frac{\delta}{2}\right)^2\right]}{e^{-\frac{1}{2}\Sigma x_i^2 - \frac{1}{2}\Sigma y_i^2}} = e^{-(m\delta^2/4)+\delta[\Sigma y_i - \Sigma x_i]}. \tag{10.5}$$

Since the distribution of $Y - X$ depends only on $\eta - \xi$, it is easily seen that condition (10.2) is satisfied.

Some further results can be obtained through extension to the sequential case of Theorems 8.1 and 9.1.

THEOREM 10.3. *Suppose that G is of the type described in Theorem 9.1, let $Y = f(X_1, X_2, \cdots)$ be maximal invariant under G, let $v(\theta)$ be maximal invariant under $\bar{G}$, and let the set of values of $v(\theta)$ corresponding to ω_0 and ω_1 be $\bar{\omega}_0$ and $\bar{\omega}_1$, respectively. If among all tests of $\bar{\omega}_0$ against $\bar{\omega}_1$ based on Y, the test φ minimizes $\sup_{v(\theta)\epsilon\bar{\omega}_0+\bar{\omega}_1} E_\theta(n)$ subject to*

$$(10.6)\qquad \begin{aligned} E_\theta\varphi(Y) &\leq \alpha \text{ if } v(\theta)\epsilon\bar{\omega}_0 \\ E_\theta\varphi(Y) &\geq \beta \text{ if } v(\theta)\epsilon\bar{\omega}_1, \end{aligned}$$

then φ also minimizes $\sup_{\omega_0+\omega_1} E_\theta(n)$ among all tests based on the X's and which satisfy (10.2).

As an example consider the problem of testing the hypothesis $\sigma \leq \sigma_0$ against the alternatives $\sigma \geq \sigma_1$ $(\sigma_0 < \sigma_1)$ when the X's are identically, independently normally distributed with unknown mean and variance. Since the problem remains invariant under a common translation of the X's we can take for Y of the theorem $Y = (X_2 - X_1, X_3 - X_1, \cdots)$. Equivalently we may take as our new sequence of variables $(Y_1, Y_2, \cdots)$ where

$$(10.7)\qquad Y_k = \frac{kX_{k+1} - (X_1 + \cdots + X_k)}{\sqrt{k(k+1)}}.$$

Then $Y_1, Y_2, \cdots$ are independently normally distributed with zero mean and the same variance as the X's. Hence the problem reduces to a type which we have already considered. The optimum test is based on

$$\sum_{i=1}^{m} Y_i^2 = \sum_{i=1}^{m+1} \left(X_i - \frac{X_1 + \cdots + X_{m+1}}{m+1}\right)^2.$$

It may be worth pointing out that Theorems 3.2, 8.3, 10.2 all are special cases of simple results in the general theory of statistical decision functions, of which the following is the prototype. (For a detailed treatment of this theory see, for example, [30]). Let $\{P_\theta\}$, $\theta\epsilon\Omega$, be the family of possible distributions of a random variable X, and let $\{\delta\}$ be a family of decision functions. The loss resulting from the use of $\delta(x)$ when P_θ is the true distribution is $W[\theta, \delta(x)]$ and the risk function associated with δ is $R_\delta(\theta) = E_\theta W[\theta, \delta(X)]$. Let λ be a probability measure over Ω, and let δ_λ be a decision function that minimizes $\int R_\delta(\theta)\, d\lambda(\theta)$. Then if λ is such that

$$(10.8)\qquad R_{\delta_\lambda}(\theta) \leqq \int R_{\delta_\lambda}(\zeta)\, d\lambda(\zeta) \text{ for all } \theta \in \Omega,$$

δ_λ minimizes $\sup_\theta R_\delta(\theta)$.

PROOF. Let δ^* be any other decision function. Then

$$\sup_\theta R_{\delta_\lambda}(\theta) \leq \int R_{\delta_\lambda}(\theta)\, d\lambda(\theta) \leqq \int R_{\delta^*}(\theta)\, d\lambda(\theta) \leqq \sup_\theta R_{\delta^*}(\theta).$$

In an analogous manner one can give an extension of Theorems 8.1, 9.1, 10.3.

11. Two sided tests considered as 3-decision problems. In a number of important special problems the hypothesis specifies the value of a real valued parameter or states that this parameter lies in a certain interval, and it is desired to test this hypothesis against the obvious two-sided class of alternatives. It seems that in nearly any problem of this kind that would arise in practice one would want to decide when rejecting the hypothesis, whether the true parameter value lies below or above the hypothetical ones. If for example one rejects the hypothesis that the means of two normal populations are equal, one usually wants to decide which of the two is larger. It would therefore seem most natural to formulate such problems as 3-decision problems.

Problems of this kind, as all problems of hypothesis testing, naturally are special cases of the general decision problem formulated by Wald. We shall here consider the case that upper bounds are given for the probabilities of certain types of errors and thereby obtain a formulation, which is closely analogous to the classical formulation of hypothesis testing discussed in this paper, and which will allow immediate application of a large portion of the theory discussed here.

Consider the case that Ω is partitioned into 3 parts, ω, ω_1, ω_2 where in a certain sense ω lies between ω_1 and ω_2. We wish to test the hypothesis $H: \theta \epsilon \omega$. When we reject the hypothesis, we shall reach either decision D_1 that $\theta \epsilon \omega_1$ or decision D_2 that $\theta \epsilon \omega_2$. Correspondingly we prescribe two positive numbers α_1, α_2 and impose the restriction that

$$(11.1) \qquad \begin{aligned} P_\theta(D_1) &\leq \alpha_1 \text{ if } \theta \epsilon \omega + \omega_2 \\ P_\theta(D_2) &\leq \alpha_2 \text{ if } \theta \epsilon \omega + \omega_1 . \end{aligned}$$

Subject to this condition it is desired to maximize

$$(11.2) \qquad \begin{aligned} &P_\theta(D_1) \text{ for } \theta \epsilon \omega_1 \\ &P_\theta(D_2) \text{ for } \theta \epsilon \omega_2 . \end{aligned}$$

A test will now consist of two non-negative functions ϕ_1 and ϕ_2 satisfying

$$(11.3) \qquad \phi_1(x) + \phi_2(x) \leq 1,$$

with the convention that when $X = x$ the decision D_i will be taken with probability $\phi_i(x)$ $(i = 1, 2)$.

There is no difficulty concerning the extension of the notions of invariance or sufficient statistic, in fact these notions obviously apply to the general decision problem. The notion of unbiasedness is extended in the obvious way by the condition

$$(11.4) \qquad \begin{aligned} P_\theta(D_1) &\geq \alpha_1 \text{ for } \theta \epsilon \omega_1 \\ P_\theta(D_2) &\geq \alpha_2 \text{ for } \theta \epsilon \omega_2 . \end{aligned}$$

One then obtains the following

THEOREM 11.1. *Suppose that for testing the hypothesis H_1: $\theta \epsilon \omega + \omega_2$ against the alternatives $\theta \epsilon \omega_1$ at level of significance α_1, the test ϕ_1 among all unbiased tests*

is uniformly most powerful in $\omega + \omega_2$ *and uniformly least powerful in* ω_1, *and that* ϕ_2 *has the analogous property for testing* $H_2 : \theta \epsilon \omega + \omega_1$ *against* $\theta \epsilon \omega_2$ *at significance level* α_2. *If* $\phi_1(x) + \phi_2(x) \leqq 1$ *for all* x, *then among all procedures satisfying* (11.1) *and* (11.4), *the procedure* (ϕ_1, ϕ_2) *uniformly maximizes the probability of a correct decision.* (If the tests ϕ_1, ϕ_2 take on only the values 0 and 1, the condition $\phi_1(x) + \phi_2(x) \leqq 1$ states that the rejection region of each of the two hypotheses is contained in the acceptance region of the other.)

As an example consider the case that $X_1, \cdots, X_n$ are independently, normally distributed with common mean ξ and variance σ^2. Suppose we wish to test the hypothesis that $\sigma_1 \leqq \sigma \leqq \sigma_2$ where σ_1 may equal σ_2. Then it follows from Theorem 11.1 that among all unbiased procedures of level (α_1, α_2), there exists one that maximizes the probability of a correct decision uniformly in ξ, σ. This is the procedure under which decision D_1 or D_2 is taken as $\Sigma(x_i - \bar{x})^2 \leqq k_1$ or $\geqq k_2$ and the hypothesis is accepted otherwise. Here the k's are determined by

$$P(\Sigma(x_i - \bar{x})^2 \leqq k_1 \mid \sigma_1) = \alpha_1 \tag{11.5}$$
$$P(\Sigma(x_i - \bar{x})^2 \geqq k_2 \mid \sigma_2) = \alpha_2 .$$

REFERENCES

[1] B. DE FINETTI, "La prévision: ses lois logiques, ses sources subjectives," *Annales de l'Institut Henri Poincaré*, Vol. 7 (1937), p. 1.

[2] P. R. HALMOS AND L. J. SAVAGE, "Applications of the Radon-Nikodym theorem to the theory of sufficient statistics," *Annals of Math. Stat.*, Vol. 20 (1949), p. 225.

[3] P. L. HSU, "Analysis of variance from the power function standpoint," *Biometrika*, Vol 32 (1941), p. 62.

[4] P. L. HSU, "On the power function of the E^2-test and the T^2-test," *Annals of Math. Stat.*, Vol. 16 (1945), p. 278

[5] G. HUNT AND C. STEIN, "Most stringent tests of statistical hypotheses," unpublished.

[6] E. L. LEHMANN, "On families of admissible tests," *Annals of Math. Stat.*, Vol. 18 (1947), p. 97.

[7] E. L. LEHMANN, "On the theory of testing composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 18 (1947), p. 473.

[8] E. L. LEHMANN, "On the totality of transformations leaving a statistical problem invariant," unpublished.

[9] E. L. LEHMANN AND H. SCHEFFÉ, "On the problem of similar regions," *Proc. Nat. Acad Sci.*, Vol. 33 (1947), p. 382.

[10] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions and unbiased estimation," unpublished

[11] E. L. LEHMANN AND C. STEIN, "Most powerful tests of composite hypotheses. I. Normal distributions," *Annals of Math. Stat*, Vol. 19 (1948), p. 495.

[12] E. L. LEHMANN AND C. STEIN, "On the theory of some non-parametric hypotheses," *Annals of Math. Stat.*, Vol 20 (1949), p 28

[13] J. NEYMAN, "Su un teorema concernente le cosidotte statistiche sufficienti," *Giorn. Ist. Ital Att.*, Vol 6 (1935), p. 320.

[14] J NEYMAN, "Sur la vérification des hypothèses statistiques composées," *Bull. Soc. Math. France*, Vol 63 (1935), p 1

[15] J NEYMAN, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans Roy. Soc. London*, Series A, Vol. 236 (1937), p. 333.

[16] J. NEYMAN AND E. S. PEARSON, "On the use and interpretation of certain test criteria for purposes of statistical inference," *Biometrika*, Vol. 20A (1928), p. 175 and p. 263.

[17] J. Neyman and E. S. Pearson, "On the testing of statistical hypotheses in relation to probability a priori," *Proc. Camb. Phil. Soc.*, Vol. 29 (1933), p. 492.

[18] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc. London*, Series A, Vol. 231 (1933), p. 289.

[19] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses. I. Unbiased critical regions of type A and type A_1," *Stat. Res. Mem.*, Vol. 1 (1936), p. 1.

[20] E. J. G. Pitman, "Tests of hypotheses concerning location and scale parameters," *Biometrika*, Vol. 31 (1939), p. 200.

[21] J. B. Simaika, "On an optimum property of two important statistical tests," *Biometrika*, Vol. 32 (1941), p. 70.

[22] C. Stein and A. Wald, "Sequential confidence intervals for the mean of a normal distribution with known variance," *Annals of Math. Stat.*, Vol. 18 (1947), p. 427.

[23] A. Wald, "On the power function of the analysis of variance test," *Annals of Math. Stat.*, Vol. 13 (1942), p. 434.

[24] A. Wald, "On the principles of statistical inference," *Notre Dame Math. Lectures*, No. 1 (1942).

[25] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Am. Math. Soc.*, Vol. 54 (1943), p. 426.

[26] A. Wald, "Statistical decision functions which minimize the maximum risk," *Annals of Math.*, Vol. 46 (1945), p. 265.

[27] A. Wald, "An essentially complete class of admissible decision functions," *Annals of Math. Stat.*, Vol. 18 (1947), p. 549.

[28] A. Wald, *Sequential analysis*, John Wiley and Sons, 1947.

[29] A. Wald, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (1947), p. 279.

[30] A. Wald, "Statistical decision functions," *Annals of Math. Stat.*, Vol. 20 (1949), p. 165.

[31] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Annals of Math. Stat.*, Vol. 19 (1948), p. 326.

[32] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Annals of Math. Stat.*, Vol. 9 (1938), p. 60.

[33] J. Wolfowitz, "Additive partition functions and a class of statistical hypotheses," *Annals of Math. Stat.*, Vol. 13 (1942), p. 247.

[34] J. Wolfowitz, "Non-parametric statistical inference," *Proceedings of the Berkeley Symposium on mathematical statistics and probability* (1949), p. 93.

[35] J. Wolfowitz, "The power of the classical tests associated with the normal distribution," *Annals of Math. Stat.*, Vol. 20 (1949), p. 540.

Some related papers not referred to in the text.

[36] T. W. Anderson, "On the theory of testing serial correlation," *Skand. Aktuarietidskrift*, (1948), p. 88.

[37] R. A. Fisher, *The design of experiments*, Oliver and Boyd, 1935.

[38] M. N. Ghosh, "On the problem of similar regions," *Sankhya*, Vol. 8 (1948), p. 329.

[39] P. G. Hoel, "Testing the homogeneity of Poisson frequencies," *Annals of Math. Stat.*, Vol. 16 (1945), p. 362.

[40] P. G. Hoel, "Discriminating between binomial distributions," *Annals of Math. Stat.*, Vol. 18 (1947), p. 556.

[41] P. G. Hoel, "On the uniqueness of similar regions," *Annals of Math. Stat.*, Vol. 19 (1948), p. 66.

[42] E. L. Lehmann, "Some comments on large sample tests," *Proceedings of the Berkeley Symposium on mathematical statistics and probability* (1949), p. 451.

[43] H. B. Mann and A. Wald, "On the choice of the number of intervals in the application of the χ^2-test," *Annals of Math. Stat.*, Vol. 13 (1942), p. 306.

[44] J. Neyman, *Lectures and conferences on mathematical statistics*, Graduate School of the U. S. Dept. of Agriculture, 1938.

[45] J. NEYMAN, "Basic ideas and some recent results of the theory of testing statistical hypotheses," *Journal Roy. Stat. Soc.*, Vol. cv (1942), p. 292.
[46] J. NEYMAN, "On a statistical problem arising in routine analysis and in sampling inspection of mass production," *Annals of Math. Stat.*, Vol 12 (1941), p. 46.
[47] S. N. ROY, "Notes on testing composite hypotheses, I, II," *Sankhya*, Vol. 8 (1947), p. 257 and Vol. 9 (1948), p. 19.
[48] H. SCHEFFÉ, "On the theory of testing composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 13 (1942), p. 280.
[49] H. SCHEFFÉ, "On the ratio of the variances of two normal samples," *Annals of Math. Stat.*, Vol. 13 (1942), p. 371.
[50.] C. STEIN, "A two sample test for a linear hypothesis whose power is independent of the variance," *Annals of Math. Stat.*, Vol. 16 (1945), p. 243.
[51] A. WALD, "Asymptotically most powerful tests of statistical hypotheses," *Annals of* Math. Stat., Vol. 12 (1941), p. 1.
[52] A WALD, "Some examples of asymptotically most powerful tests," *Annals of Math. Stat.*, Vol. 12 (1941), p. 396.
[53] A. WALD, "On the efficient design of statistical investigations," *Annals of Math. Stat.*, Vol. 14 (1943), p. 134.

SAMPLE CRITERIA FOR TESTING OUTLYING OBSERVATIONS[1]

By Frank E. Grubbs

University of Michigan and Ballistic Research Laboratories

1. Summary. The problem of testing outlying observations, although an old one, is of considerable importance in applied statistics. Many and various types of significance tests have been proposed by statisticians interested in this field of application. In this connection, we bring out in the Historical Comments notable advances toward a clear formulation of the problem and important points which should be considered in attempting a complete solution. In Section 4 we state some of the situations the experimental statistician will very likely encounter in practice, these considerations being based on experience. For testing the significance of the largest observation in a sample of size n from a normal population, we propose the statistic

$$\frac{S_n^2}{S^2} = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where $x_1 \leq x_2 \leq \cdots \leq x_n$, $\bar{x}_n = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i$ *and* $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$.

A similar statistic, S_1^2/S^2, can be used for testing whether the smallest observation is too low.

It turns out that

$$\frac{S_n^2}{S^2} = 1 - \frac{1}{n-1}\left(\frac{x_n - \bar{x}}{s}\right)^2 = 1 - \frac{1}{n-1} T_n^2,$$

where $s^2 = \frac{1}{n} \Sigma(x_i - \bar{x})^2$, and T_n is the studentized extreme deviation already suggested by E. Pearson and C. Chandra Sekar [1] for testing the significance of the largest observation. Based on previous work by W. R. Thompson [12], Pearson and Chandra Sekar were able to obtain certain percentage points of T_n without deriving the exact distribution of T_n. The exact distribution of S_n^2/S^2 (or T_n) is apparently derived for the first time by the present author.

For testing whether the two largest observations are too large we propose the statistic

$$\frac{S_{n-1,n}^2}{S^2} = \frac{\sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1,n})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \qquad \bar{x}_{n-1,n} = \frac{1}{n-2} \sum_{i=1}^{n-2} x_i$$

---

[1] This paper has been extracted from a thesis approved for the Degree of PhD at the University of Michigan.

and a similar statistic, $S_{1,2}^2/S^2$, can be used to test the significance of the two smallest observations. The probability distributions of the above sample statistics

TABLE I

Table of Percentage Points for $\frac{S_n^2}{S^2}$ *or* $\frac{S_1^2}{S^2}$

Percentage Points

n	1%	2.5%	5%	10%
3	.0001	.0007	.0027	.0109
4	.0100	.0248	.0494	.0975
5	.0442	.0808	.1270	.1984
6	.0928	.1453	.2032	.2826
7	.1447	.2066	.2696	.3503
8	.1948	.2616	.3261	.4050
9	.2411	.3101	.3742	.4502
10	.2831	.3526	.4154	.4881
11	.3211	.3901	.4511	.5204
12	.3554	.4232	.4822	.5483
13	.3864	.4528	.5097	.5727
14	.4145	.4792	.5340	.5942
15	.4401	.5030	.5559	.6134
16	.4634	.5246	.5755	.6306
17	.4848	.5442	.5933	.6461
18	.5044	.5621	.6095	.6601
19	.5225	.5785	.6243	.6730
20	.5393	.5937	.6379	.6848
21	.5548	.6076	.6504	.6958
22	.5692	.6206	.6621	.7058
23	.5827	.6327	.6728	.7151
24	.5953	.6439	.6829	.7238
25	.6071	.6544	.6923	.7319

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{where} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$S_n^2 = \sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2 \quad \text{where} \quad \bar{x}_n = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i$$

$$S_1^2 = \sum_{i=2}^{n} (x_i - \bar{x}_1)^2 \quad \text{where} \quad \bar{x}_1 = \frac{1}{n-1} \sum_{i=2}^{n} x_i$$

are derived for a normal parent and tables of appropriate percentage points are given in this paper (Table I and Table V). Although the efficiencies of the above tests have not been completely investigated under various models for outlying

observations, it is apparent that the proposed sample criteria have considerable intuitive appeal. In deriving the distributions of the sample statistics for testing the largest (or smallest) or the two largest (or two smallest) observations, it was first necessary to derive the distribution of the difference between the extreme observation and the sample mean in terms of the population σ. This probability

TABLE IA

Table of Percentage Points for $T_n = \dfrac{x_n - \bar{x}}{s}$ *or* $T_1 = \dfrac{\bar{x} - x_1}{s}$

n	1%	2.5%	5%	10%
3	1.414	1.414	1.412	1.406
4	1.723	1.710	1.689	1.645.
5	1.955	1.917	1.869	1.791
6	2.130	2.067	1.996	1.894
7	2.265	2.182	2.093	1.974
8	2.374	2.273	2 172	2.041
9	2.464	2.349	2.237	2.097
10	2.540	2.414	2.294	2.146
11	2.606	2.470	2.343	2.190
12	2.663	2.519	2.387	2.229
13	2.714	2.562	2.426	2.264
14	2.759	2.602	2.461	2.297
15	2.800	2.638	2.493	2.326
16	2.837	2.670	2.523	2.354
17	2.871	2.701	2.551	2.380
18	2.903	2.728	2.577	2.404
19	2.932	2.754	2.600	2.426
20	2.959	2.778	2.623	2.447
21	2.984	2.801	2.644	2.467
22	3.008	2.823	2.664	2.486
23	3.030	2.843	2.683	2.504
24	3.051	2.862	2.701	2.520
25	3.071	2.880	2.717	2.537

$$x_1 \leq x_2 \leq x_3 \cdots \leq x_n$$

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

distribution was apparently derived first by A. T. McKay [11] who employed the method of characteristic functions. The author was not aware of the work of McKay when the simplified derivation for the distribution of $\dfrac{x_n - \bar{x}}{\sigma}$ outlined in Section 5 below was worked out by him in the spring of 1945, McKay's result

being called to his attention by C. C. Craig. It has been noted also that K. R. Nair [20] worked out independently and published the same derivation of the distribution of the extreme minus the mean arrived at by the present author see *Biometrika*, Vol. 35, May, 1948. We nevertheless include part of this derivation in Section 5 below as it was basic to the work in connection with the derivations given in Sections 8 and 9. Our table is considerably more extensive than Nair's table of the probability integral of the extreme deviation from the sample mean in normal samples, since Nair's table runs from $n = 2$ to $n = 9$, whereas our Table II is for $n = 2$ to $n = 25$. The present work is concluded with some examples.

2. Introduction. Scientific data are collected usually for purposes of interpretation and if proper use is to be made of the information thus obtained then some decision should be reached or some action taken as a result of analyzing the data. In many cases a critical examination of the data collected is necessary in order to insure that the results of sampling are representative of the thing or process we are examining. Quite frequently our observations do not appear to be consistent with one another, i.e. the data may seem to display non-homogeneities and the group of observations as a whole may not appear to represent a random sample from, say, a single normal population or universe. In particular, one or more of the observations may have the appearance of being "outliers" and we are interested here in determining once and for all whether such observations should be retained in the sample for interpreting results or whether they should be regarded as being inconsistent with the remaining observations. It is clear that rejection of the "outliers" in a sample will in a great number of cases lead to a different course of action than would have been taken had such observations been retained in the sample. Actually, the rejection of "outlying" observations may be just as much a practical (or common sense) problem as a statistical one and sometimes the practical or experimental viewpoint may naturally outweigh any statistical contributions. In this connection, the concluding remarks of Rider's survey [2] are pertinent: "In the final analysis it would seem that the question of the rejection or the retention of a discordant observation reduces to a question of common sense. Certainly the judgment of an experienced observer should be allowed considerable influence in reaching a decision. This judgment can undoubtedly be aided by the application of one or more tests based on the theory of probability, but any test which requires an inordinate amount of calculation seems hardly to be worth while, and the testimony of any criterion which is based upon a complicated hypothesis should be accepted with extreme caution." Hence, it would appear that statistical tests of significance for judging or testing "outliers" come into importance either in supporting doubtful practical viewpoints or in providing a basis for action in the absence of sufficient experimental knowledge of underlying causes in an investigation. Indeed, the latter two situations are met quite frequently in practice.

In the present treatment, we intend to throw some light beyond the work

TABLE II

Probability Integral of the Extreme Minus the Mean, u_n, in Normal Samples of n Observations (Pop. S.D. as unit) $P(u_n \leq u)$

u \ n	2	3	4	5	6	7	8	9	u
.00	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00
.05	.05637	.00309	.00017	.00001	.00000	.00000	.00000	.00000	.05
.10	.11246	.01231	.00134	.00015	.00002	.00000	.00000	.00000	.10
.15	.16800	.02745	.00445	.00072	.00012	.00002	.00000	.00000	.15
.20	.22270	.04817	.01033	.00221	.00047	.00010	00002	.00000	.20
.25	.27633	.07403	.01966	.00520	.00137	.00036	.00010	.00003	.25
.30	.32863	.10450	.03292	.01033	.00324	00101	.00032	.00010	.30
.35	.37938	.13896	.05040	.01820	.00656	.00236	.00085	.00031	.35
.40	.42839	.17677	.07218	.02935	.01191	.00482	.00195	00079	.40
.45	.47548	.21724	.09816	.04416	.01982	.00880	.00398	.00178	.45
.50	.52050	.25968	.12807	.06288	.03080	.01507	.00737	.00360	.50
.55	.56332	.30344	.16152	.08559	.04525	.02390	.01261	.00665	.55
.60	.60386	.34788	.19801	.11219	.06344	.03583	.02022	.01140	.60
.65	.64203	.39243	.23697	.14246	.08547	.05121	03067	.01836	.65
.70	.67780	.43656	.27781	.17602	.11130	.07030	.04437	02800	.70
.75	.71116	.47983	.31992	.21242	.14076	.09318	.06164	.04076	.75
.80	.74210	.52185	.36274	.25113	.17353	.11978	.08263	.05698	.80
85	.77067	.56230	.40571	.29160	.20920	.14993	.10739	.07688	.85
.90	.79691	.60095	.44835	.33325	.24727	.18329	.13578	.10055	.90
.95	.82089	63761	.49021	.37555	.28721	.21945	.16757	.12791	.95
1.00	.84270	.67214	.53093	.41795	.32847	.25791	.20240	.15877	1.00
1.05	.86244	.70448	.57020	.45999	.37050	.29815	23980	.19280	1.05
1.10	.88021	.73459	.60777	.50125	.41276	.33961	.27927	.22957	1.10
1.15	.89612	.76248	.64346	.54136	.45478	.38173	.32025	.26858	1.15
1.20	.91031	.78817	.67713	.58001	.49611	.42401	.36220	.30931	1.20
1.25	.92290	.81174	.70870	.61697	.53638	.46595	.40457	.35117	1.25
1.30	.93401	.83325	.73812	.65205	.57525	.50712	.44685	.39362	1.30
1.35	.94376	.85280	.76540	.68513	.61249	.54716	.48857	.43613	1.35
1.40	.95229	.87049	.79055	.71612	.64788	.58574	.52933	.47822	1.40
1.45	.95970	.88644	.81364	.74497	.68129	.62263	.56878	.51945	1.45
1.50	.96611	.90075	.83472	.77170	.71261	65762	.60663	.55944	1.50
1.55	.97162	.91355	.85390	.79632	.74180	.69058	.64265	.59789	1.55
1.60	.97635	.92495	.87127	.81890	.76885	.72143	.67668	.63456	1.60
1.65	.98038	.93506	.88693	.83949	.79378	.75013	.70862	.66925	1.65
1.70	.98379	.94400	.90099	.85820	.81664	.77666	.73839	.70184	1.70
1.75	.98667	.95187	.91358	.87513	.83750	.80107	.76597	.73225	1.75
1.80	.98909	.95877	.92480	.89037	.85646	.82341	.79139	.76046	1.80
1.85	.99111	.96480	.93476	90405	.87360	.84376	.81469	.78647	1.85
1 90	.99279	.97005	.94358	.91628	.88903	.86220	.83593	.81032	1.90
1.95	.99418	.97461	.95135	.92716	.90288	.87885	.85522	.83207	1.95

TABLE II—*Continued*

u \ n	2	3	4	5	6	7	8	9	u
2.00	.99532	.97854	.95818	.93682	.91526	.89381	.87264	.85183	2.00
2.05	.99626	.98193	.96416	.94536	.92627	90721	.88832	.86968	2.05
2.10	.99702	98483	.96938	.95289	.93605	.91916	90236	88574	2.10
2.15	.99764	98731	.97392	.95949	.94468	.92977	91490	90012	2 15
2 20	99814	.98942	.97785	.96527	.95229	93917	92604	91296	2 20
2.25	.99854	.99121	.98125	.97032	.95897	94746	.93591	.92438	2 25
2.30	.99886	.99273	.98418	.97470	.96482	95476	94462	93448	2.30
2 35	.99911	99400	.98669	.97850	.96992	.96114	95229	.94340	2.35
2.40	.99931	.99507	.98883	.98178	.97435	.96672	.95900	.95125	2.40
2.45	.99947	99596	.99066	.98461	97819	.97158	.96487	.95812	2.45
2 50	.99959	.99670	.99222	98703	.98151	.97580	.96999	.96412	2.50
2.55	.99969	.99732	.99353	.98911	.98436	97944	.97443	96935	2.55
2.60	.99976	.99782	.99464	99088	.98681	.98259	.97827	.97389	2.60
2 65	.99982	.99824	.99557	.99238	.98891	.98529	.98158	.97781	2.65
2.70	.99987	.99858	.99635	.99365	.99070	.98761	.98443	98120	2.70
2 75	99990	.99886	.99701	.99473	.99223	.98959	98688	98411	2.75
2 80	.99992	.99909	.99755	.99564	.99352	.99128	.98897	.98661	2 80
2.85	.99994	.99928	.99800	.99640	.99461	.99272	99075	98874	2.85
2.90	.99996	.99943	.99838	.99704	.99553	.99393	.99227	.99056	2.90
2 95	.99997	.99955	.99868	.99757	.99631	.99496	.99355	.99211	2.95
3 00	.99998	.99964	.99894	.99801	.99696	.99582	.99464	.99342	3.00
3 05	99998	.99972	.99914	.99838	.99750	.99655	.99556	.99453	3.05
3.10	.99999	.99978	.99931	.99868	.99795	.99716	.99632	.99546	3.10
3.15	.99999	.99983	.99945	.99893	.99832	.99766	.99697	.99625	3.15
3.20	.99999	.99987	.99956	.99913	.99863	.99808	.99750	.99690	3.20
3.25	1.00000	99990	.99965	.99930	.99889	.99843	.99795	.99745	3.25
3 30		99992	.99972	.99944	.99910	.99872	.99832	.99791	3.30
3 35		99994	.99978	.99955	.99927	.99896	.99863	.99829	3.35
3.40		.99995	99983	.99964	.99941	.99916	.99889	.99860	3.40
3.45		.99996	.99986	.99971	.99953	.99932	.99910	.99886	3.45
3.50		.99997	.99989	.99977	.99962	.99945	.99927	.99908	3.50
3.55		.99998	.99992	.99982	.99970	.99956	.99941	.99925	3.55
3.60		.99998	.99994	.99986	.99976	.99965	.99952	.99940	3.60
3 65		.99999	.99995	.99989	.99981	.99972	.99962	.99951	3.65
3.70		99999	.99996	.99991	.99985	.99977	.99969	.99961	3.70
3.75		.99999	.99997	.99993	.99988	.99982	.99976	.99969	3.75
3 80		1.00000	99998	.99995	.99991	.99986	.99981	.99975	3.80
3 85			.99998	99996	.99993	.99989	.99985	.99980	3.85
3.90			.99999	.99997	.99994	.99991	.99988	.99984	3.90
3 95			.99999	99997	.99995	.99993	.99990	.99987	3.95

TABLE II—*Continued*

u \ n	2	3	4	5	6	7	8	9	u
4.00			.99999	.99998	.99996	.99995	.99992	.99990	4.00
4.05			.99999	.99999	.99997	.99996	.99994	99992	4.05
4.10			1 00000	.99999	.99998	.99997	.99995	.99994	4.10
4.15				99999	.99998	99997	.99996	.99995	4.15
4.20				.99999	.99999	.99998	.99997	.99996	4.20
4.25				.99999	.99999	.99998	99998	.99997	4.25
4.30				1.00000	.99999	.99999	.99998	.99998	4.30
4.35					.99999	.99999	.99999	.99998	4.35
4.40					1.00000	.99999	.99999	.99999	4.40
4.45						99999	.99999	.99999	4.45
4.50						1.00000	.99999	.99999	4.50
4.55							1 00000	.99999	4.55
4.60								1.00000	4.60

u \ n	10	11	12	13	14	15	16	17	u
.25	.00001	.00000	.00000	00000	00000	.00000	.00000	.00000	25
.30	.00003	00001	.00000	.00000	.00000	.00000	.00000	.00000	.30
.35	.00011	.00004	.00001	00001	.00000	.00000	.00000	.00000	.35
.40	.00032	.00013	.00005	.00002	.00001	.00000	.00000	.00000	.40
.45	.00080	.00036	.00016	.00007	.00003	.00001	.00001	.00000	.45
.50	.00176	.00086	.00042	.00021	.00010	.00005	.00002	.00001	.50
.55	.00351	.00185	.00098	.00051	.00027	.00014	.00008	.00004	.55
.60	.00643	.00363	.00204	.00115	.00065	.00037	.00021	00012	.60
.65	01098	.00657	00393	.00235	00141	.00084	.00050	.00030	.65
.70	.01766	01113	00702	.00443	.00279	.00176	.00111	.00070	.70
.75	02694	.01780	01177	.00777	.00514	.00339	.00224	.00148	75
.80	.03928	.02707	.01865	.01285	.00886	.00610	00420	00289	.80
.85	.05503	.03938	02818	.02016	.01442	.01031	.00738	.00527	.85
.90	.07444	.05510	04077	.03017	.02232	.01652	.01222	.00904	.90
.95	.09761	.07448	.05682	.04334	.03305	.02521	01922	.01460	.95
1.00	.12452	.09763	.07655	06000	.04703	.03687	.02889	.02265	1.00
1.05	15497	.12454	.10008	.08041	.06460	.05190	.04169	.03348	1.05
1.10	.18867	.15503	.12737	.10464	.08595	.07060	.05799	.04762	1.10
1.15	.22520	.18879	.15825	.13263	.11116	.09315	.07806	.06541	1.15
1.20	.26407	.22542	.19240	.16420	.14013	11957	.10203	.08706	1.20
1 25	.30475	.26442	.22941	.19901	.17263	.14973	.12987	.11264	1.25
1.30	34666	.30525	26876	.23662	.20830	.18336	.16140	.14207	1.30
1.35	.38924	.34734	30992	.27650	.24667	22005	.19629	.17509	1.35
1.40	.43196	.39011	35229	31810	.28721	.25931	.23411	21135	1.40
1.45	.47430	.43302	39529	.36082	.32934	.30058	.27433	.25036	1.45

TABLE II—*Continued*

w \ n	10	11	12	13	14	15	16	17	w
1 50	.51583	47555	.43838	40408	.37244	.34327	.31636	.29156	1.50
1.55	.55615	.51726	.48104	.44733	.41595	.38676	.35960	.33434	1.55
1.60	.59495	.55774	52282	.49004	.45930	.43046	.40342	.37807	1.60
1.65	63196	.59668	.56332	.53178	.50199	.47384	.44726	.42216	1.65
1.70	.66699	.63380	.60221	.57216	.54358	.51641	.49058	.46602	1.70
1.75	69991	.66892	.63925	.61086	.58370	.55773	.53289	.50915	1.75
1.80	.73063	.70189	.67424	.64763	.62204	.59744	.57380	.55108	1.80
1.85	.75912	.73264	.70704	.68229	.65838	.63528	.61297	.59144	1.85
1.90	.78538	.76113	.73758	.71472	.69254	.67102	.65016	.62992	1.90
1.95	.80945	.78737	.76584	.74486	.72443	.70453	.68516	.66630	1.95
2.00	.83141	.81140	.79183	77269	.75399	73571	.71786	.70042	2.00
2.05	.85133	83330	.81560	.79824	78121	.76453	.74819	.73218	2.05
2.10	86932	.85314	.83721	.82155	.80614	.79101	.77614	76153	2.10
2.15	.88550	87105	.85678	84271	.82885	.81519	.80174	.78849	2.15
2 20	89998	.88713	.87440	.86183	.84941	.83715	82505	.81311	2.20
2.25	.91290	.90151	89021	.87002	.86795	.85699	.84616	.83545	2.25
2.30	.92437	.91431	.90432	.89441	.88458	.87484	.86518	.85563	2.30
2.35	.93453	.92568	.91688	.90812	.89943	.89081	88224	.87375	2.35
2.40	.94348	.93572	.02799	.92030	.91264	.90504	.89748	.88997	2.40
2.45	95134	.94457	.93781	.93106	.92435	.91766	.91101	.90440	2.45
2.50	95823	.95233	94644	.94055	.93468	.92883	.92300	.91720	2.50
2.55	96424	.95912	.95400	94887	.94376	.93866	.93357	.92850	2.55
2.60	.96948	.96504	.96060	95616	.95172	.94728	.94285	.93844	2.60
2.65	97401	.97019	96635	.96251	.95866	.95482	˙05098	.94715	2.65
2.70	.97793	.97464	.97134	.96802	.96471	.96139	.05807	.95475	2.70
2.75	98131	.97849	97565	.97280	.96995	.96709	.96423	.90137	2.75
2 80	.98422	.98180	.97937	.97693	.97448	.97203	.96957	.96712	2.80
2.85	98671	.98464	.98257	.98048	97839	.97629	.97418	.97208	2.85
2.90	.98883	.98708	.98531	.98353	.98174	.97995	.97816	.97636	2.90
2.95	.99064	.98915	.98765	.98614	.98462	.98300	.98156	.98003	2.95
3.00	.99218	.99092	.98965	.98837	98708	98578	.98448	.98318	3.00
3.05	.99348	.99242	.99134	.99026	.98917	.98807	.98697	.98587	3.05
3.10	.99458	.99369	.99278	.99187	.99095	.99002	.98909	.98816	3.10
3.15	.99551	.99476	99400	99323	.99245	.99167	.99089	.99010	3.15
3.20	.99628	.99566	.99502	.99437	.99372	.99307	.99241	.99175	3.20
3.25	.99694	.99641	99588	.99534	.99479	99424	.99369	.99314	3.25
3 30	.99748	.99704	99660	99615	.99569	.99523	99477	.99431	3.30
3.35	.99793	.99757	.99720	.99682	.99644	.99606	.99568	.99529	3.35
3.40	.99831	.99801	.99770	.99739	.99707	.99676	.99644	.99611	3.40
3.45	.99862	99837	.99812	.99786	.99760	.99733	.99707	.99680	3.45

TABLE II—*Continued*

u \ n	10	11	12	13	14	15	16	17	u
3.50	.99888	.99867	.99846	.99825	.99803	.99781	.99759	.99737	3.50
3.55	.99909	.99892	.99875	.99857	.99839	.99821	.99803	.99785	3.55
3.60	.99926	.99912	.99898	.99884	.99869	99854	.99839	.99824	3.60
3.65	.99940	.99929	.99917	.99906	.99894	.99881	.99869	.99857	3.65
3.70	.99952	.99943	.99933	.99924	.99914	.99904	.99894	.99883	3.70
3.75	.99961	.99954	.99946	.99938	.99930	.99922	.99914	.99905	3.75
3.80	.99969	.99963	.99957	.99950	.99944	.99937	.99930	.99923	3.80
3.85	.99975	.99970	99965	.99960	.99955	.99949	.99944	.99938	3.85
3.90	.99980	.99976	.99972	.99968	.99964	.99959	.99955	.99950	3.90
3.95	.99984	.99981	.99978	.99974	.99971	.99967	.99964	.99960	3.95
4.00	.99988	.99985	.99982	99980	.99977	.99974	.99971	.99968	4.00
4.05	.99990	.99988	.99986	99984	.99982	.99979	.99977	.99974	4.05
4 10	.99992	.99991	.99989	.99987	.99985	.99983	.99981	.99979	4.10
4.15	.99994	.99993	.99991	.99990	.99988	.99987	.99985	.99984	4.15
4.20	.99995	99994	.99993	.99992	.99991	.99990	.99988	.99987	4.20
4.25	99996	.99995	.99995	.99994	.99993	.99992	.99991	.99990	4.25
4.30	.99997	.99996	.99996	.99995	.99994	.99993	.99993	.99992	4.30
4.35	.99998	.99997	99997	99996	.99996	.99995	.99994	.99993	4.35
4.40	.99998	.99998	.99997	.99997	.99996	.99996	.99995	.99995	4.40
4.45	.99999	.99998	.99998	.99998	.99997	.99997	.99996	.99996	4.45
4 50	.99999	.99999	99998	99998	99998	.99998	.99997	.99997	4.50
4.55	99999	.99999	99999	.99999	.99998	99998	.99998	.99997	4.55
4.60	.99999	99999	.99999	.99999	99999	.99998	.99998	.99998	4.60
4.65	1.00000	.99999	99999	.99999	.99999	.99999	.99999	.99998	4.65
4.70		1.00000	.99999	99999	.99999	.99999	.99999	.99999	4.70
4 75			1.00000	1.00000	99999	.99999	.99999	.99999	4.75
4.80					1.00000	.99999	.99999	.99999	4.80
4.85						1 00000	1.00000	1.00000	4.85

u \ n	18	19	20	21	22	23	24	25	u
.50	.00001	.00000	.0000	.0000	0000	.0000	.0000	.0000	.50
55	.00002	.00001	.0000	.0000	.0000	.0000	.0000	.0000	.55
.60	.00007	.00004	.0000	.0000	.0000	.0000	.0000	.0000	.60
.65	.00018	.00011	.0001	.0000	0000	.0000	.0000	.0000	.65
.70	.00044	.00028	.0002	.0001	.0001	.0000	.0000	.0000	.70
.75	00098	.00065	.0004	.0003	0002	.0001	.0001	.0001	.75
80	.00199	.00137	.0009	.0007	0004	.0003	.0002	.0001	.80
.85	.00377	00270	.0019	.0014	0010	.0007	.0005	.0004	.85
.90	.00669	00494	0037	.0027	.0020	.0015	.0011	.0008	.90
.95	.01118	.00853	0065	.0049	.0038	.0029	.0022	.0017	.95

TABLE II—*Continued*

u \ n	18	19	20	21	22	23	24	25	u
1 00	.01775	.01391	.0109	.0085	.0067	.0052	.0041	.0032	1.00
1.05	.02690	.02161	.0174	.0139	.0112	.0090	.0072	.0058	1.05
1.10	03911	.03212	.0264	.0217	.0178	.0146	.0120	.0099	1.10
1.15	.05481	.04592	.0385	.0322	.0270	.0226	.0190	.0159	1.15
1.20	.07428	.06338	.0541	.0461	.0394	.0336	.0287	.0244	1.20
1.25	.09769	.08472	.0735	.0637	.0553	.0479	.0416	.0360	1.25
1.30	.12504	.11005	.0969	.0853	.0750	.0660	.0581	.0512	1.30
1.35	.15618	.13930	.1242	.1108	.0988	.0882	.0786	.0701	1.35
1.40	.19080	.17225	.1555	.1404	.1267	.1144	.1033	.0932	1.40
1.45	.22848	.20851	.1903	.1736	.1585	.1446	.1320	.1204	1.45
1.50	.26869	.24761	.2282	.2103	.1938	.1786	.1646	.1516	1.50
1.55	.31084	28899	.2687	.2498	.2322	.2159	.2007	.1866	1.55
1.60	.35430	33202	.3111	.2916	.2732	.2560	.2399	.2248	1.60
1.65	.39845	37607	.3549	.3349	.3162	.2984	.2816	.2658	1.65
1.70	.44269	.42052	.3994	.3794	.3604	.3424	.3252	.3089	1.70
1.75	.48645	46476	.4440	.4242	.4053	.3872	.3699	.3534	1.75
1.80	.52924	.50827	4881	.4687	.4502	.4323	.4152	.3987	1.80
1.85	.57065	.55058	.5312	.5125	.4945	.4771	.4603	.4441	1.85
1.90	.61031	59130	.5729	.5549	.5377	.5209	.5047	.4890	1.90
1.95	64796	.63011	.6127	.5958	.5794	.5634	.5479	.5328	1.95
2 00	.68340	.66678	.6506	.6348	.6193	.6042	.5895	.5752	2.00
2.05	71650	.70114	.6861	.6714	.6570	.6429	.6291	.6156	2.05
2.10	.74719	.73311	.7193	.7058	.6924	.6793	.6665	.6540	2.10
2.15	.77545	.76262	7500	.7375	.7254	.7133	.7015	.6899	2.15
2.20	.80132	.78971	.7782	7670	.7558	.7448	.7340	.7234	2.20
2.25	.82486	.81440	8041	7938	.7838	.7738	.7640	.7543	2.25
2 30	84616	.83679	.8275	.8184	.8093	.8003	.7914	.7827	2.30
2.35	.86533	.85699	8487	.8405	.8324	.8244	.8164	.8085	2.35
2.40	.88251	.87511	.8678	.8605	.8533	.8461	.8390	.8319	2.40
2.45	.89783	89129	8848	.8784	.8720	.8656	.8593	.8530	2.45
2.50	.91142	.90568	.9000	.8943	.8887	.8831	.8775	.8719	2.50
2.55	92345	.91842	.9134	.9084	.9035	.8985	.8936	.8888	2.55
2 60	.93404	.92965	.9253	.9209	.9166	.9123	.9080	.9037	2.60
2.65	.94332	.93951	.9357	.9319	.9282	.9244	.9207	.9169	2.65
2.70	.95144	.94814	.9448	.9416	.9382	.9351	.9318	.9286	2.70
2.75	.95852	.95567	.9528	9500	.9472	.9444	.9415	.9387	2.75
2.80	.96466	.96220	.9598	.9573	.9549	.9524	.9500	.9476	2.80
2 85	96997	.96787	9658	.9637	.9616	9595	.9574	.9553	2.85
2.90	.97456	.97275	9710	.9692	.9674	.9656	.9638	.9620	2.90
2.95	.97850	.97696	.9754	.9739	.9724	.9709	.9693	.9678	2.95

TABLE II—*Continued*

u \ n	18	19	20	21	22	23	24	25	u
3.00	98187	.98057	.9793	.9780	.9767	.9753	.9741	.9728	3.00
3.05	.98476	98365	.9825	.9814	.9803	.9793	.9781	.9771	3.05
3.10	.98722	.98629	.9853	.9844	9835	.9826	9816	.9807	3.10
3.15	.98931	.98852	.9877	.9869	9862	.9853	.9846	.9838	3.15
3.20	.99108	.99042	.9898	.9891	9884	.9878	.9871	9865	3.20
3.25	.99258	.99202	.9915	.9909	.9904	.9898	.9893	.9887	3.25
3.30	.99384	.99337	.9929	.9924	.9920	.9915	.9911	.9906	3.30
3.35	.99490	.99451	.9941	.9937	.9933	.9930	9926	.9922	3.35
3.40	.99579	.99546	.9951	.9948	.9945	.9942	9939	.9936	3.40
3.45	99653	.99626	.9960	9957	.9955	.9952	9949	.9947	3.45
3.50	.99715	.99693	.9967	.9965	.9963	.9961	.9958	.9956	3.50
3.55	.99766	.99748	.9973	.9971	.9969	.9968	.9966	.9964	3.55
3.60	.99809	.99794	.9978	.9976	.9975	9973	.9972	.9971	3.60
3.65	.99844	99832	.9982	.9981	.9979	.9978	9977	.9976	3.65
3.70	.99873	.99863	.9985	.9984	.9983	.9982	.9982	.9981	3.70
3.75	.99897	.99889	.9988	.9987	.9986	.9986	.9985	9984	3.75
3.80	.99917	.99910	.9990	.9990	.9989	.9988	.9988	.9988	3.80
3.85	.99933	.99927	.9992	.9992	.9991	.9991	.9990	.9990	3.85
3.90	.99946	.99941	.9994	.9993	.9993	.9993	.9992	.9992	3.90
3.95	.99956	.99953	.9995	.9995	.9994	.9994	.9994	.9994	3.95
4.00	.99965	.99962	.9996	.9996	.9995	.9995	.9995	.9995	4.00
4.05	.99972	.99969	9997	.9996	.9996	.9996	.9996	.9996	4.05
4.10	.99977	.99975	.9997	9997	.9997	.9997	.9997	.9997	4.10
4 15	.99982	99980	.9998	.9998	.9998	.9998	.9998	.9998	4.15
4.20	.99986	.99984	.9998	.9998	.9998	.9998	.9998	.9998	4.20
4.25	.99989	99987	.9999	.9999	.9999	.9999	.9999	.9999	4.25
4.30	.99991	.99990	.9999	.9999	.9999	.9999	.9999	.9999	4.30
4.35	.99993	.99992	.9999	.9999	.9999	.9999	.9999	.9999	4.35
4.40	.99994	.99994	.9999	.9999	.9999	.9999	.9999	.9999	4.40
4.45	.99995	99995	1.0000	.9999	.9999	.9999	.9999	.9999	4.45
4.50	.99996	.99996		1.0000	1.0000	1.0000	.9999	.9999	4.50
4.55	.99997	.99997					1.0000	1 0000	4.55
4.60	.99998	.99997							4.60
4.65	.99998	.99998							4.65
4.70	.99998	.99998							4.70
4.75	.99999	.99998							4.75
4.80	.99999	.99999							4.80
4 85	.99999	.99999							4.85
4.90	1 00000	1.00000							4.90

that has already been done [1], [2], [3], [4], [11], [12], [20] on the problem of testing outlying observations statistically and to see just where our contributions fit into this corner of mathematical statistics. First, however, we give a very brief history of the problem.

3. Historical comments. A survey of statistical literature indicates that the problem of testing the significance of outlying observations received considerable attention prior to 1937. Since this date, however, published literature on the subject seems to have been unusually scant—perhaps because of inherent difficulties in the problem as pointed out by E. S. Pearson and C. Chandra Sekar [1]. These authors made some important contributions to the problem of outlying observations by bringing clearly into the foreground the concept of efficiency of tests which may be used in view of admissible alternative hypotheses.

In 1933, P. R. Rider [2] published a rather comprehensive survey of work on the problem of testing the significance of outlying observations up to that date. The test criteria surveyed by Rider appear to impose as an initial condition that the standard deviation, σ, of the population from which the items were drawn should be known accurately. In connection with such tests requiring accurate knowledge of σ, we mention (1) Irwin's criteria [3] which utilize the difference between the first two individuals or the difference between the second and third individuals in random samples from a normal population and (2) the range[2] or maximum dispersion [4], [5], [6], [7], [8], [9], [10], [18] of a sample which has been advocated by "Student" [4] and others for testing the significance of outlying observations. We remark further that a natural statistic to use for testing an "outlier" is the difference between such an extreme observation and the sample mean. In 1935, McKay [11] published a note on the distribution of the last-mentioned statistic and by means of a rather elaborate procedure obtained a recurrence relation between the distribution of the extreme minus the mean in samples of n from a normal universe and the distribution of this statistic in samples of $n - 1$ from the same parent. McKay gave also an approximate expression for the upper percentage points of the distribution but did not tabulate the exact distribution due to the complicity of the multiple integrals involved. McKay pointed out that if K_p denotes the p-th semi-invariant of the distribution of $x_n - \bar{x}$ (where x_n is the largest observation) and K'_p refers similarly to the distribution of x_n, then $K_1 = K'_1 - \mu$, $K_2 = K'_2 - \frac{1}{n}$ and $K_p = K'_p$ ($p \geq 3$ where $\mu = E(x_i)$. Nair [20] has tabulated the distribution of the difference between the extreme and sample mean for $n = 2$ to $n = 9$.

Under certain circumstances, accurate knowledge concerning σ may be available as, for example, in using "daily control" tests [4], [18] the population standard deviation may be estimated in some cases with sufficient precision from past

[2] The derivation for the exact distribution of the range is given in reference [9], 1942; however, Dr. L. S. Dederick of the Ballistic Research Laboratory also derived the exact distribution of the range in an unpublished Aberdeen Proving Ground Report (1926).

data. In general, however, an accurate estimate of σ may not be available and it becomes necessary to estimate the population standard deviation from the single sample involved or "Studentize" [18], [20] the statistic to be used, thus providing a true measure of the risks involved in the significance test advocated for testing outlying observations. W. R. Thompson [12] apparently had this very point in mind when he devised an exact test in his paper, "On a Criterion for the Rejection of Observations and the Distribution of the Ratio of the Deviation to the Sample Standard Deviation," which appeared in 1935. Thompson showed that if

$$T_i = \frac{x_i - \bar{x}}{s}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$ and x_i is an observation selected arbitrarily from a random sample of n items drawn from a normal parent, then the probability density function of

$$t = \frac{T\sqrt{n-2}}{\sqrt{n-1-T^2}}$$

is given by "Student's" t-distribution with $f = n - 2$ degrees of freedom.

Pearson and Chandra Sekar have given a rather comprehensive study of Thompson's criterion in an interesting and important paper [1] which appeared in 1936. They discussed also some very important viewpoints which should be taken into consideration when dealing with the problem of testing outlying observations. By setting up alternatives to the null-hypothesis H_0 that all items in the sample come from the same population, Pearson and Chandra Sekar point out that if only one of the observations actually came from a population with divergent mean, then Thompson's criterion would be very useful, whereas if two or more of the observations are truly outlying then the criterion $| x_i - \bar{x} | \geq T_0 s$ may be quite ineffective, particularly if the sample contains less than about 30 or 40 observations

A point of major interest concerning Thompson's work nevertheless is that he proposed an *exact* test for the hypothesis that all of the observations came from the same normal population. With regard to the use of an arbitrary observation in Thompson's test, however, it should be borne in mind that the problem of finding the probability that an arbitrary observation will be outlying is different from that of finding the probability that a particular observation (the largest, for example) will be outlying with respect to the other $n - 1$ observations of the sample.

As a final point concerning the paper of Pearson and Chandra Sekar [1], we see that for the n values of T_i arranged in order of magnitude taking account of sign, say

$$T^{(1)}, T^{(2)}, \cdots T^{(n)},$$

then

$$T^{(1)} \geq T^{(2)} \geq T^{(3)} \cdots \geq T^{(n)}.$$

The above authors show that the form of the total distribution of all the T_i at its extremes depend only on $T^{(1)}$ and $T^{(n)}$. This is because for some combinations of sample size and percentage points the algebraic upper limit for $T^{(2)}$ and algebraic lower limit for $T^{(n-1)}$ do not extend into the "tails" of the total distribution. Hence, the following probability law holds for $T^{(1)}$ when $T^{(1)} \geq$ the algebraic maximum of $T^{(2)}$:

$$p\{T^{(1)}\} = Np(T).$$

Likewise,

$$p\{T^{(n)}\} = Np(T)$$

for $T^{(n)} \leq$ algebraic minimum of $T^{(n-1)}$. Therefore, Pearson and Chandra Sekar were able to use Thompson's table [12] and give (for some sample sizes) upper probability limits for $T^{(1)} = \dfrac{x_i - \bar{x}}{s}$ for the highest observation and lower probability limits for $T^{(n)} = \dfrac{x_i - \bar{x}}{s}$ for the lowest observation without actually obtaining the exact probability distribution of $T^{(1)}$ and $T^{(n)}$. Hence, the appearance of the table of percentage points on page 318 of their paper [1] was a substantial contribution to the problem of testing outlying observations since an exact test for the significance of a single outlying observation was provided for the case where an accurate estimate of σ is not available. (The exact distribution of $T^{(1)}$ or $T^{(n)}$ is derived later in this work.)

With the above highlights of historical background in mind, we turn now to a consideration of the types of problems the experimenter may be faced with in testing "outlying" observations.

4. Statement of hypotheses in tests of outliers. Once the sample results of an experiment are available, the practicing statistician may be confronted with one or more of the following distinct situations as regards discordant observations: (a) To begin with, a very frequent or perhaps prevalent situation is that either the greatest observation or the least observation in a sample may have the appearance of belonging to a different population than the one from which the remaining observations were drawn. Here we are confronted with tests for a single outlying observation. (b) Then again, both the largest and the smallest observations may appear to be "different" from the remaining items in the sample. Here we are interested in testing the hypothesis that both the largest and the smallest observations are truly "outliers." (c) Another frequent situation is that either the two largest or the two smallest observations may have the appearance of being discordant Here we are interested in reaching a decision as to whether we should reject the two largest or the two smallest observations as not being representative of the thing we are sampling.

As to why the discordant observations in a sample may be outliers, this may be due to errors of measurement in which case we would naturally want to reject or at least "correct" such observations. On the other hand, it may be that the population we are sampling is not homogeneous in the uni-modal sense and it will consequently be desirable to know this so that we may carry out further development work on our product if possible or desirable.

Although there may be many models for outliers, we believe that an important practical case involves the situation where all the observations in the sample may be subject to the same standard error, whereas it may happen that the largest or smallest observations result from shifts in level. For example, if one observation appears unusually high compared to the others in the sample we may want to consider the hypothesis that all the observations come from a normal parent with mean μ and standard deviation σ as against the alternative hypothesis that the largest observation comes from a normal population with mean $\mu + \lambda\sigma$ ($\lambda > 0$) and standard deviation σ, whereas the remaining observations are from $N(\mu, \sigma)$.

Another case involves the situation where the largest and/or smallest observations may be from $N(\mu, \lambda\sigma)$, $\lambda > 1$, whereas the remaining observations of the sample are from the normal parent $N(\mu, \sigma)$

Although we have not investigated the power of the tests proposed herein for various models, it is believed that the exact test of Section 8 for the largest (or smallest) observation and the test of Section 9 for the two largest (or two smallest) observations possess considerable intuitive appeal for the practical situations described above.[3]

5. Distribution of the difference between the extreme and mean in samples of n from a normal population. The simultaneous density function of n independent observations from a normal parent with zero mean and variance σ^2 which are arranged in order of magnitude is given by

$$(1) \qquad dF(x_1, x_2, \cdots, x_n) = \frac{n!}{(\sqrt{2\pi}\,\sigma)^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right] \quad dx_1\, dx_2 \cdots dx_n$$

subject to $x_1 \leq x_2 \leq \cdots \leq x_n$

Since

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{n}{n-1}(x_n - \bar{x})^2 + \sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2$$

where

$$\bar{x}_n = \frac{1}{n-1}\sum_{i=1}^{n-1} x_i,$$

[3] The author is indebted to J. W. Tukey and S. S. Wilks for calling attention to an incorrect distribution function in the originally submitted manuscript on which several yet-to-be proved or disproved statements concerning optimum properties of statistics in this paper were based.

then

$$\sum_{i=1}^{n} x_i^2 = n\bar{x}^2 + \frac{n}{n-1}(x_n - \bar{x})^2 + \frac{n-1}{n-2}(x_{n-1} - \bar{x}_n)^2$$

$$(2) \qquad + \frac{n-2}{n-3}(x_{n-2} - \bar{x}_{n,n-1})^2 + \cdots + \frac{3}{2}\left(x_3 - \frac{x_1 + x_2 + x_3}{3}\right)^2 + \frac{2}{1}\left(x_2 - \frac{x_1 + x_2}{2}\right)^2$$

where

$$\bar{x}_{n,n-1} = \frac{1}{n-2}\sum_{i=1}^{n-2} x_i, \quad \text{etc.}$$

and consequently we find that we are particularly interested in the following Helmert orthogonal transformation:

$$(3) \qquad \begin{aligned} \sqrt{2\cdot 1}\,\sigma\eta_2 &= -x_1 + x_2, \\ \sqrt{3\cdot 2}\,\sigma\eta_3 &= -x_1 - x_2 + 2x_3, \\ &\;\vdots \\ \sqrt{n(n-1)}\,\sigma\eta_n &= -x_1 - x_2 - x_3 - x_4 - \cdots - x_r \\ &\qquad - \cdots - x_{n-1} + (n-1)x_n, \\ \sqrt{n}\,\sigma\eta_{n+1} &= x_1 + x_2 + x_3 + x_4 + \cdots + x_r + \cdots + x_{n-1} + x_n. \end{aligned}$$

The above transformation will lead to the distribution of the difference between the extreme and sample mean in terms of the unknown population σ for samples of n from a normal parent. Since, however, K. R. Nair (*Biometrika*, May, 1948) has already published the details independently, we will only record here for later reference that the density function of $\eta_2, \eta_3, \cdots, \eta_n$ (after integrating η_{n+1} over $-\infty \leq \eta_{n+1} \leq +\infty$) is

$$(4) \qquad dF(\eta_2, \eta_3, \cdots \eta_n) = \frac{n!}{(\sqrt{2\pi})^{n-1}} \exp\left[-\frac{1}{2}\sum_{i=2}^{n} \eta_i^2\right] d\eta_2\, d\eta_3 \cdots d\eta_n$$

where the η_i are restricted by the relations

$$(5) \qquad \infty \geq \eta_2 \geq 0, \qquad \sqrt{\frac{r}{r-2}}\,\eta_r \geq \eta_{r-1}.$$

Upon making the transformations

$$(6) \qquad \frac{\sqrt{r(r-1)}}{r}\,\eta_r = \frac{x_r - \bar{x}}{\sigma} = u_r, \qquad (r = 2, 3, \cdots, n),$$

defining

$$(7) \qquad F_n(u) = \int_0^u dF(u_n) = \text{probability } u_n \leq u,$$

and integrating the u_n over their appropriate ranges we find the cumulative probability integrals of the extreme deviation from the sample mean (in terms of the population σ) for $n = 2, 3, \cdots$ to be

$$F_2(u) = 2\sqrt{2}\int_0^u \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(2x^2)}\,dx = \frac{2}{\sqrt{\pi}}\int_0^u e^{-x^2}\,dx,$$

a well-known result, where for $n = 2$, x is either the sample standard deviation, the difference between the extreme and sample mean, the mean deviation or the semi-range.

$$F_3(u) = \frac{3\sqrt{3}}{\sqrt{2}}\int_0^u \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{3}{2}x^2)} F_2(\tfrac{3}{2}x)\,dx,$$

$$\vdots \tag{8}$$

$$F_n(u) = \frac{n\sqrt{n}}{\sqrt{n-1}}\int_0^u \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}((n)/(n-1)x^2)} F_{n-1}\left(\frac{n}{n-1}\,x\right)dx.$$

This is equivalent to the result of McKay (11), although the derivation indicated is a considerably simpler one.

Now $F_{n-1}(u)$ increases from 0 to 1 as u increases from 0 to ∞. Hence, if $F_{n-1}\left(\frac{n}{n-1}\,u\right)$ is practically unity, i.e. for $\frac{n}{n-1}\,u$ numerically large, the upper percentage points of u_n may be approximated by the normal integral

$$\begin{aligned}\int_{u_n}^{\infty} dF(u_n) &= \frac{n}{\sqrt{2\pi}}\int_{u_n}^{\infty}\exp\left[-\frac{1}{2}\frac{n}{n-1}\,u_n^2\right]\frac{\sqrt{n}}{\sqrt{n-1}}\,du_n \\ &= \frac{n}{\sqrt{2\pi}}\int_{\sqrt{n/(n-1)}\,u_n}^{\infty}\exp\left[-\frac{t^2}{2}\right]dt\end{aligned} \tag{9}$$

Formula (9) was found to be particularly useful in checking the higher probabilities in Table II.

The cumulative distribution functions (8) may be put into another form by setting

$$u_r = \frac{1}{r}\,v_r\,; \qquad r = 2, 3, \cdots, n\ .$$

Then $F_n(u)$ becomes

$$\begin{aligned}F_n(u) = \frac{\sqrt{n}}{(\sqrt{2\pi})^{n-1}}\int_0^{nu}\int_0^{v_n}\int_0^{v_{n-1}}\cdots\int_0^{v_4}\int_0^{v_3} \\ \cdot\exp\left[-\frac{1}{2}\sum_{i=2}^{n}\frac{v_i^2}{i(i-1)}\right]dv_2\,dv_3\cdots dv_n.\end{aligned} \tag{10}$$

Define the following functions:

$$H_1(x) = 1,$$

$$H_2(x) = \sqrt{2} \int_0^x \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} \cdot \frac{t^2}{2 \cdot 1}\right] H_1(t)\, dt,$$

$$\vdots$$

$$H_n(x) = \sqrt{\frac{n}{n-1}} \int_0^x \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} \cdot \frac{t^2}{n(n-1)}\right] H_{n-1}(t)\, dt.$$

Hence, the probability that the difference between the extreme and the mean in samples of n from a normal population is less than $u\sigma$ is given by the alternative forms

$$P\{u_n \leq u\sigma\} = F_n(u) = H_n(nu).$$

Of course, $H_n(nu) \to 1$ as $u \to \infty$ for any given n.

In the November 1945 issue of *Biometrika*, Godwin [13] arrived at a series of functions closely related to the $H_r(x)$ in connection with the distribution of the mean deviation in samples of n from a normal parent. In Godwin's work, he defines functions $G_r(x)$ which are related to the $H_r(x)$ by the equation

$$(2\pi)^{r/2} H_{r+1}(x) = G_r(x).$$

The $G_r(x)$ functions were computed by H. O. Hartley [15] for $r = 2, 3, \cdots 9$ only. Computations on the functions $F_n(u)$, i.e. (8), were well under way by the author before Godwin's article on the mean deviation appeared. The $H_r(x)$ or $G_r(x)$ can be used to obtain both the distribution of the difference between the extreme and mean and also the probability integral of the mean deviation. Indeed, it is believed that these functions may have a useful place in tabulating distributions of order statistics.

6. Tabulation of the distribution function, $F_n(u)$.

The tabulation of the $F_n(u)$ with ordinary computing equipment is quite laborious. However, a table model computing machine was used initially to obtain the $F_n(u)$ for $n = 2$ to $n = 15$ using formulae (8) and a numerical quadrature process.

In view of the possible general usefulness of the $H_r(x)$, these functions were also computed as a sample problem on a high-speed computing device, the ENIAC (Electronic Numerical Integrator and Computor) of the Ballistic Re-

[4] The author suggested the problem of tabulating the functions $F_n(u)$ or $H_n(nu)$ to the Computing Laboratory of the Ballistic Research Laboratories in the fall of 1945; however, due to problems of higher priority, these functions were not computed on the ENIAC until March, 1948.

search Laboratories of the Ordnance Department.[4] In this connection, the $H_r(u)$ have been computed for $r = 2$ to $r = 25$ at the Ballistic Research Laboratories. For $n = 2$, the functions $H_r(x)$ were computed to nine decimal places of accuracy on the ENIAC and at $n = 25$ about five decimal places of accuracy were obtained. In Table II we have tabulated $F_n(u)$ or $H_n(nu)$, i.e. the probability integral of the extreme minus the mean, at intervals of $u = .05\sigma$. Values computed on the table model computing machine agreed to five decimal places at $n = 15$ with values from the ENIAC. Percentage Points of the distribution are given in Table III and the moment constants may be found in Table IV. Moment constants for $n = 60$, 100, 200, 500 and 1000 were obtained by use of McKay's formulae [11] (which relate the semi-invariants of $x_n - \bar{x}$ with those of x_n) and Tippetts moments [5] for the largest observation x_n.

TABLE III

Percentage Points for Extreme Minus Mean

n	90%	95%	99%	99.5%
2	1.163	1.386	1.821	1.985
3	1.497	1.738	2.215	2.396
4	1.696	1.941	2.431	2.618
5	1.835	2.080	2.574	2.764
6	1.939	2.184	2.679	2.870
7	2.022	2.267	2.761	2.952
8	2.091	2.334	2.828	3.019
9	2.150	2.392	2.884	3.074
10	2.200	2.441	2.931	3.122
11	2.245	2.484	2.973	3.163
12	2.284	2.523	3.010	3.199
13	2.320	2.557	3.043	3.232
14	2.352	2.589	3.072	3.261
15	2.382	2.617	3.099	3.287
16	2.409	2.644	3.124	3.312
17	2.434	2.668	3.147	3.334
18	2.458	2.691	3.168	3.355
19	2.480	2.712	3.188	3.375
20	2.500	2.732	3.207	3.393
21	2.519	2.750	3.224	3.409
22	2.538	2.768	3.240	3.425
23	2.555	2.784	3.255	3.439
24	2.571	2.800	3.269	3.453
25	2.587	2.815	3.282	3.465

TABLE IV

Moment Constants for Extreme Minus Mean

n	Mean	Std. Dev.	α_3	α_4
2	.5642	.4263	.9953	3.8692
3	.8463	.4755	.8296	3.7135
4	1.0294	.4916	.7675	3.6717
5	1.1630	.4974	.7372	3.6560
6	1.2672	.4993	.7165	3.6511
7	1.3522	.4991	.7042	3.6503
8	1.4236	.4979	.6959	3.6518
9	1.4850	.4962	.6900	3.6546
10	1.5388	.4943	.6857	3.6582
11	1.5864	.4923	.6827	3.6622
12	1.6292	.4902	.6804	3.6663
13	1.6680	.4881	.6788	3.6705
14	1.7034	.4861	.6777	3.6746
15	1.7359	.4841	.6770	3.6787
20	1.867	.475	.677	3.700
60	2.319	.436	.699	3.801
100	2.508	.418	.712	3.855
200	2.746	.395	.737	3.932
500	3.037	.368	.771	4.033
1000	3.241	.350	.794	4.105

7. Relation between the distribution of the largest minus the mean of all n observations and the largest minus the mean of the remaining n-1 items. The following relation is of interest concerning these two statistics:

Let

$$u_n = x_n - \frac{x_1 + x_2 + \cdots + x_n}{n}$$
$$= \frac{1}{n} \{(n - 1)\, x_n - x_1 - x_2 - \cdots - x_{n-1}\}.$$

Let

$$v_n = x_n - \frac{x_1 + x_2 + \cdots + x_{n-1}}{n - 1}$$
$$= \frac{1}{n - 1} \{(n - 1)\, x_n - x_1 - x_2 - \cdots - x_{n-1}\}.$$

Hence,

$$v_n = \frac{n}{n - 1}\, u_n$$

or

$$P(v_n \leq t_0) = P\left(\frac{n}{n-1} u_n \leq t_0\right) = P\left\{u_n \leq \frac{n-1}{n} t_0\right\},$$

i.e. the probability integral of the largest minus the mean of the other observations may be obtained by interpolation on the distribution of the largest minus the mean of all n items in the sample.

8. The distribution of S_n^2/S^2 and S_1^2/S^2. As indicated in the Summary, we proposed the sample criterion

$$\frac{S_n^2}{S^2} = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \leq k, \qquad \bar{x}_n = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i,$$

for testing the significance of the largest observation and the criterion

$$\frac{S_1^2}{S^2} = \frac{\sum_{i=2}^{n} (x_i - \bar{x}_1)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \leq k, \qquad \bar{x}_1 = \frac{1}{n-1} \sum_{i=2}^{n} x_i,$$

for testing whether the smallest observation is outlying. We now find the probability distribution of S_n^2/S^2; hence, also that of S_1^2/S^2.

Returning to the density function

$$dF(\eta_2, \eta_3, \cdots, \eta_n) = \frac{n!}{(\sqrt{2\pi})^{n-1}} \exp\left[-\frac{1}{2} \sum_{i=2}^{n} \eta_i^2\right] d\eta_2 \, d\eta_3 \cdots d\eta_n$$

of Section 5, we make the polar transformation

$$\begin{aligned} \eta_2 &= r \sin\theta_n \sin\theta_{n-1} \cdots \sin\theta_4 \sin\theta_3, \\ \eta_3 &= r \sin\theta_n \sin\theta_{n-1} \cdots \sin\theta_4 \cos\theta_3, \\ \eta_4 &= r \sin\theta_n \sin\theta_{n-1} \cdots \cos\theta_4, \\ &\ \,\vdots \\ \eta_{n-1} &= r \sin\theta_n \cos\theta_{n-1}, \\ \eta_n &= r \cos\theta_n. \end{aligned} \tag{11}$$

Now

$$\sum_{i=2}^{n} \eta_i^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 = r^2$$

and

$$\sum_{i=2}^{n-1} \eta_i^2 = \sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2 = r^2 \sin^2\theta_n.$$

Hence,

$$\sin^2 \theta_n = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

The Jacobian of the above transformation is

$$r^{n-2} \sin^{n-3}\theta_n \sin^{n-4}\theta_{n-1} \cdots \sin^3 \theta_6 \sin^2 \theta_5 \sin \theta_4,$$

and since $0 \leq r \leq \infty$

$$\begin{aligned} &dF(\theta_n, \theta_{n-1}, \cdots, \theta_5, \theta_4, \theta_3) \\ (12) \quad &= \frac{n!}{(2\pi)^{(n-1)/2}} 2^{(n-3)/2} \Gamma\left(\frac{n-1}{2}\right) \sin^{n-3} \theta_n \cdots \sin^2 \theta_5 \sin \theta_4 \, d\theta_n \cdots d\theta_5 \, d\theta_4 \, d\theta_3. \end{aligned}$$

Since the restrictions on the η_i are

$$\eta_2 \geq 0, \qquad \sqrt{\frac{r}{r-2}}\,\eta_r \geq \eta_r - 1, \qquad r \geq 3,$$

we have

$$\tan \theta_n \cos \theta_{n-1} = \frac{\eta_{n-1}}{\eta_n}, \qquad n \geq 4,$$

or

$$\tan \theta_n \leq \sqrt{\frac{n}{n-2}} \sec \theta_{n-1}, \qquad n \geq 4,$$

and

$$0 \leq \theta_3 \leq \frac{\pi}{3}.$$

Thus, letting $K_n = \frac{n!}{(2\pi)^{(n-1)/2}} 2^{(n-3)/2} \Gamma\left(\frac{n-1}{2}\right)$, we see that

$$(13) \quad K_n \int_0^{\pi/3} \int_0^{l_3} \cdots \int_0^{l_{n-2}} \int_0^{l_{n-1}} \sin^{n-3} \theta_n \cdots \sin^2 \theta_5 \sin \theta_4 \, d\theta_n \cdots d\theta_4 \, d\theta_3 = 1,$$

where $l_r = \tan^{-1} \sqrt{\frac{r+1}{r-1}} \sec \theta_r$.

Upon reversing the order of integration (the variable limits are monotonic) we get for $n = 3$

$$K_3 \int_0^{\pi/3} d\theta_3 = 1,$$

so that

$$(14) \qquad P(\theta_3 \leq \theta) = K_3 \int_0^{\theta} d\theta_3 \qquad 0 \leq \theta \leq M_3 = \tan^{-1} \sqrt{3 \cdot 1}.$$

When $n = 4$, we obtain

$$K_4 \int_0^{m_4} \int_0^{\pi/3} \sin \theta_4 \, d\theta_3 \, d\theta_4 + K_4 \int_{m_4}^{M_4} \int_{L_4}^{\pi/3} \sin \theta_4 \, d\theta_3 \, d\theta_4 = 1$$

where

$$m_r = \tan^{-1} \sqrt{\frac{r}{r-2}}, M_r = \tan^{-1} \sqrt{r(r-2)} \text{ and } L_r = \sec^{-1} \sqrt{\frac{r-2}{r}} \tan \theta_r,$$

so that

(15a) $$P(\theta_4 \leq \theta) = \frac{K_4}{K_3} \int_0^{\theta} \sin \theta_4 \, d\theta_4 \qquad \text{when } 0 \leq \theta \leq m_4 = \tan^{-1} \sqrt{\frac{4}{2}}$$

and

(15b) $$P(\theta_4 \leq \theta) = \frac{K_4}{K_3} \int_0^{m_4} \sin \theta_4 \, d\theta_4 + K_4 \int_{m_4}^{\theta} \int_{L_4}^{\pi/3} \sin \theta_4 \, d\theta_3 \, d\theta_4$$

$$\text{when } m_4 = \tan^{-1} \sqrt{\frac{4}{2}} \leq \theta \leq M_4 = \tan^{-1} \sqrt{4 \cdot 2}.$$

When $n = 5$, we get,

$$K_5 \int_0^{m_5} \int_0^{m_4} \int_0^{\pi/3} \sin^2 \theta_5 \sin \theta_4 \, d\theta_3 \, d\theta_4 \, d\theta_5 + K_5 \int_0^{m_5} \int_{m_4}^{M_4} \int_{L_4}^{\pi/3} \sin^2 \theta_5 \sin \theta_4 \, d\theta_3 \, d\theta_4 \, d\theta_5$$

$$+ K_5 \int_{m_5}^{M_5} \int_{L_5}^{M_4} \int_{L_4}^{\pi/3} \sin^2 \theta_5 \sin \theta_4 \, d\theta_3 \, d\theta_4 \, d\theta_5 = 1$$

$\Big($where $L_4 = \sec^{-1} \sqrt{\frac{2}{4}} \tan \theta_4$ is to be taken as 0 whenever $\theta_4 \leq m_4 = \tan^{-1} \sqrt{\frac{4}{2}}\Big)$ so that

(16a) $$P(\theta_5 \leq \theta) = \frac{K_5}{K_4} \int_0^{\theta} \sin^2 \theta_5 \, d\theta_5 \qquad \text{when } 0 \leq \theta \leq m_5 = \tan^{-1} \sqrt{\frac{5}{3}}$$

and

(16b) $$P(\theta_5 \leq \theta) = \frac{K_5}{K_4} \int_0^{m_5} \sin^2 \theta_5 \, d\theta_5 + K_5 \int_{m_5}^{\theta} \int_{L_5}^{M_4} \int_{L_4}^{\pi/3} \sin^2 \theta_5 \sin \theta_4 \, d\theta_3 \, d\theta_4 \, d\theta_5$$

where $m_5 = \tan^{-1} \sqrt{\frac{5}{3}} \leq \theta \leq M_5 = \tan^{-1} \sqrt{5 \cdot 3}$,

and we put $L_4 = \sec^{-1} \sqrt{\frac{2}{4}} \tan \theta_4 = 0$ whenever $\theta_4 \leq m_4 = \tan^{-1} \sqrt{\frac{4}{2}}$.

For a sample of n items

$$(17a)\quad \begin{aligned} P(\theta_n \leq \theta) &= \frac{n}{\sqrt{\pi}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \int_0^\theta \sin^{n-3}\theta_n \, d\theta_n \\ &= \frac{n}{2} I_{\sin^2\theta}\left(\frac{n-2}{2}, \frac{1}{2}\right) \qquad \text{when } 0 \leq \theta \leq \tan^{-1}\sqrt{\frac{n}{n-2}}^{5} \end{aligned}$$

and

$$(17b)\quad \begin{aligned} P(\theta_n \leq \theta) &= \frac{n}{2} I_{n/(2(n-1))}\left(\frac{n-2}{2}, \frac{1}{2}\right) \\ &+ K_n \int_{m_n}^{\theta} \int_{L_n}^{M_{n-1}} \int_{L_{n-1}}^{M_{n-2}} \cdots \int_{L_4}^{\pi/2} \sin^{n-3}\theta_n \cdots \sin\theta_4 \, d\theta_3 \, d\theta_4 \cdots d\theta_n \end{aligned}$$

for

$$m_n = \tan^{-1}\sqrt{\frac{n}{n-2}} \leq \theta \leq M_n = \tan^{-1}\sqrt{n(n-2)}$$

where $I_x(p, q)$ is K. Pearson's Incomplete Beta Function Ratio [19]. It is to be understood in (17) that

$$L_i = \sec^{-1}\sqrt{\frac{i-2}{i}} \tan\theta_i \quad \text{for} \quad i = 4, 5, \cdots, n-1$$

is to be taken as zero when $\theta_i \leq \tan^{-1}\sqrt{\frac{i}{i-2}}$.

Percentage points for the sample statistic

$$\sin^2\theta_n = \frac{S_n^2}{S^2} = \frac{\sum_{i=1}^{n-1}(x_i - \bar{x}_n)^2}{\sum_{i=1}^{n}(x_i - \bar{x})}$$

or the statistic S_1^2/S_2 are given in Table I and were obtained by inverse interpolation on the tabulation of the probability integral (17) above. Percentage points for the Pearson and Chandra Sekar statistics, $T_n = \frac{x_n - \bar{x}}{s}$ or $T_1 = \frac{\bar{x} - x_1}{s}$ (where $s^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$), are given in Table IA. The statistics S_n^2/S^2 and T_n are related by the formula

$$\frac{S_n^2}{S^2} = 1 - \frac{T_n^2}{n-1}.$$

[5] It has been noted that (17a) gives a good approximation to (17b) when $\theta \geq \tan\sqrt{\frac{n}{n-2}}$ provided we are interested in the important practical region $P \leq .10$, at least for $n \leq 25$.

The statistic T_n (or T_1) is easier to compute than S_n^2/S^2 (or S_1^2/S^2). The tabulation of the multiple integral (17) was carried out on the Bell Relay Computors at the Ballistic Research Laboratories.

9. **The distribution of $S^2_{n-1,n}/S^2$ and $S^2_{1,2}/S^2$.** As indicated in the Summary, the proposed criterion for judging the significance of the two largest observations is

$$S^2_{n-1,n}/S^2 = \frac{\sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1,n})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \leq k \qquad \text{where } \bar{x}_{n-1,n} = \frac{1}{n-2}\sum_{i=1}^{n-2} x_i,$$

and that for testing the two smallest observations is

$$S^2_{1,2}/S^2 = \frac{\sum_{i=3}^{n} (x_i - \bar{x}_{1,2})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \leq k \qquad \text{where } \bar{x}_{1,2} = \frac{1}{n-2}\sum_{i=3}^{n} x_i.$$

From the preceding section, we note that

$$\sum_{i=2}^{n} \eta_i^2 = r^2, \qquad \sum_{i=2}^{n-2} \eta_i^2 = r^2 \sin^2 \theta_n \sin^2 \theta_{n-1}.$$

Hence,

$$\sin^2\theta_n \sin^2\theta_{n-1} = \frac{\sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1,n})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \tag{18}$$

so that if we find the distribution of

$$\sin^2 \theta_n \sin^2 \theta_{n-1} = \sin^2 \Delta_n \text{, say,}$$

then we have the distribution of $S^2_{n-1,n}/S^2$ and hence also that of $S^2_{1,2}/S^2$, i.e.

$$P\{\sin^2 \Delta_n \leq k\} = P\{\Delta_n \leq \sin^{-1} \sqrt{k}\}. \tag{19}$$

Returning to the multiple integral (13), let

$$\sin \Delta_n = \sin \theta_n \sin \theta_{n-1},$$

$$\Delta_i = \theta_i, \qquad 3 \leq i \leq n-1.$$

The Jacobian of this transformation is given by

$$\frac{\partial(\theta_n, \cdots, \theta_3)}{\partial(\Delta_n, \cdots, \Delta_3)} = \frac{\cos \Delta_n}{\sqrt{\sin^2 \Delta_{n-1} - \sin^2 \Delta_n}}.$$

The limits of integration for Δ_n are given by

$$0 \leq \Delta_n \leq \sin^{-1} \frac{\sqrt{n} \sin \Delta_{n-1}}{\sqrt{2(n-1) - (n-2)\sin^2 \Delta_{n-1}}}$$

and, of course, those for $\Delta_{n-1}, \cdots, \Delta_3$ are the same as the limits for $\theta_{n-1}, \cdots, \theta_3$ respectively. Hence, substituting in (13), we obtain

$$K_n \int_0^{\pi/3} \int_0^{\tan^{-1}\sqrt{4/2}\sec\Delta_3} \cdots \int_0^{\tan^{-1}\sqrt{n-1/n-3}\sec\Delta_{n-2}} \int_0^{\sin^{-1}\frac{\sqrt{n}\sin\Delta_{n-1}}{\sqrt{2(n-1)-(n-2)\sin^2\Delta_{n-1}}}} \tag{20}$$

$$\frac{\sin^{n-3}\Delta_n \sin^{n-4}\Delta_{n-1}\cdots\sin^2\Delta_5 \sin\Delta_4 \cos\Delta_n \, d\Delta_n \cdots d\Delta_3}{\sin^{n-3}\Delta_{n-1}\sqrt{\sin^2\Delta_{n-1} - \sin^2\Delta_n}} = 1.$$

Reversing the order of integration, we have

$$K_n \int_0^{\sin^{-1}\sqrt{\frac{n(n-3)}{(n-1)(n-2)}}} \int_{\sin^{-1}\frac{\sqrt{2(n-1)}\sin\Delta_n}{\sqrt{n+(n-2)\sin^2\Delta_n}}}^{\tan^{-1}\sqrt{(n-1)(n-3)}} \int_{\sec^{-1}\sqrt{\frac{n-3}{n-1}}\tan\Delta_{n-1}}^{\tan^{-1}\sqrt{(n-2)(n-4)}} \tag{21}$$

$$\cdots \int_{\sec^{-1}\sqrt{2/4}\tan\Delta_4}^{\pi/3} \frac{\sin^{n-3}\Delta_n \sin^{n-4}\Delta_{n-1}\cdots\sin\Delta_4\cos\Delta_n \, d\Delta_3 \cdots d\Delta_n}{\sin^{n-3}\Delta_{n-1}\sqrt{\sin^2\Delta_{n-1} - \sin^2\Delta_n}} = 1$$

$\Big($for $\Delta_i \leq \tan^{-1}\sqrt{\frac{i}{i-2}}$, then $\sec^{-1}\sqrt{\frac{i-2}{i}}\tan\Delta_i$ is to be put equal to zero where $i \geq 4\Big)$ so that for $n = 4$,

$$P(\Delta_4 \leq \Delta) = K_4 \int_0^{\Delta} \int_{\sin^{-1}\frac{\sqrt{6}\sin\Delta_4}{\sqrt{2+\sin^2\Delta_4}}}^{\pi/3} \frac{\sin\Delta_4\cos\Delta_4 \, d\Delta_3 \, d\Delta_4}{\sin\Delta_3\sqrt{\sin^2\Delta_3 - \sin^2\Delta_4}} \tag{22}$$

where $0 \leq \Delta \leq \sin^{-1}\sqrt{\frac{2}{3}}$,

and for $n = 5$,

$$P(\Delta_5 \leq \Delta) = K_5 \int_0^{\Delta} \int_{\sin^{-1}\frac{\sqrt{4\cdot 2}\sin\Delta_5}{\sqrt{5+3\sin^2\Delta_5}}}^{\tan^{-1}\sqrt{4\cdot 2}} \tag{23}$$

$$\cdot \int_{\sin^{-1}\sqrt{2/4}\tan\Delta_4}^{\pi/3} \frac{\sin^2\Delta_5\cos\Delta_5 \, d\Delta_3 \, d\Delta_4 \, d\Delta_5}{\sin\Delta_4\sqrt{\sin^2\Delta_4 - \sin^2\Delta_5}}$$

where $0 \leq \Delta \leq \sin^{-1}\sqrt{\frac{5}{6}}$, etc.

We remark that an obvious extension of the above principles should lead to the distributions of

$$S^2_{n-2,n-1,n}/S^2 \quad \text{and} \quad S^2_{1,2,3}/S^2,$$

$$S^2_{n-3,n-2,n-1,n}/S^2 \quad \text{and} \quad S^2_{1,2,3,4}/S^2,$$

etc. although the tabulation of such probability integrals may be exceedingly difficult.

The problem of tabulating the probability integral (21) involves a double quadrature process and has been carried out on the Bell Relay Computors at the Ballistic Research Laboratories for $n = 4$ to $n = 20$, inclusive. Table V gives some useful percentage points for these sample sizes.

TABLE V

Table of Percentage Points for $\dfrac{S^2_{n-1,n}}{S^2}$ *or* $\dfrac{S^2_{1,2}}{S^2}$

n	1%	2.5%	5%	10%
4	.0000	.0002	.0008	.0031
5	.0035	.0090	.0183	.0376
6	.0186	.0349	.0565	.0921
7	.0440	.0708	.1020	.1479
8	.0750	.1101	.1478	.1994
9	.1082	.1492	.1909	.2454
10	.1415	.1865	.2305	.2863
11	.1736	.2212	.2666	.3226
12	.2044	.2536	.2996	.3552
13	.2333	.2836	.3295	.3843
14	.2605	.3112	.3568	.4106
15	.2859	.3367	.3818	.4345
16	.3098	.3603	.4048	.4562
17	.3321	.3822	.4259	.4761
18	.3530	.4025	.4455	.4944
19	.3725	.4214	4630	.5113
20	.3909	.4391	.4804	.5269

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{where} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S^2_{n-1,n} = \sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1,n})^2 \quad \text{where} \quad \bar{x}_{n-1,n} = \frac{1}{n-2}\sum_{i=1}^{n-2} x_i$$

$$S^2_{1,2} = \sum_{i=3}^{n} (x_i - \bar{x}_{1,2})^2 \quad \text{where} \quad \bar{x}_{1,2} = \frac{1}{n-2}\sum_{i=3}^{n} x_i$$

10. Comment on the distribution of $S^2_{1,n}/S^2$. In connection with the distribution of the statistic

$$\frac{S^2_{1,n}}{S^2} = \frac{\sum_{i=2}^{n-1} (x_i - \bar{x}_{1,2})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad \text{where} \quad \bar{x}_{1,n} = \frac{1}{n-2}\sum_{i=2}^{n-1} x_i,$$

for testing simultaneously whether the smallest and largest observations are outlying, an investigation indicates that since

$$\sum x_i^2 = n\bar{x}^{-2} + \frac{n}{n-1}(x_n - \bar{x})^2 + \frac{n-1}{n-2}(x_1 - \bar{x}_n)^2 + \frac{n-2}{n-3}(x_{n-1} - \bar{x}_{1,n})^2$$
$$+ \cdots + \frac{3}{2}\left(x_4 - \frac{x_2 + x_3 + x_4}{3}\right)^2 + 2\left(x_3 - \frac{x_2 + x_3}{2}\right)^2$$

then the transformation

$$
\begin{aligned}
\sqrt{2\cdot 1}v_2 &= -x_2 + x_3,\\
\sqrt{3\cdot 2}v_3 &= -x_2 - x_3 + 2x_4,\\
\sqrt{4\cdot 3}v_4 &= -x_2 - x_3 - x_4 + 3x_5,\\
&\;\;\vdots \qquad\qquad\qquad\qquad\qquad\qquad (24)\\
\sqrt{(n-2)(n-3)}v_{n-2} &= -x_2 - x_3 - \cdots - x_{n-2} + (n-3)x_{n-1},\\
\sqrt{(n-1)(n-2)}v_{n-1} &= -(n-2)x_1 + x_2 + x_3 + \cdots + x_{n-1},\\
\sqrt{n(n-1)}v_n &= -x_1 - x_2 - x_3 - \cdots - x_{n-1} + (n-1)x_n,\\
\sqrt{n}v_{n+1} &= x_1 + x_2 + \cdots + x_n,
\end{aligned}
$$

followed by transformations of the type (11) and that of Section 9 may lead to the distribution of $S_{1,n}^2/S^2$. However, the limits of integration do not turn out to be functions of single variables and the task of computing the resulting multiple integral may be rather difficult.

11. Examples on testing outlying observations for rejection. We now turn to the problem of applying our theory to particular practical examples of data which appear to have outlying observations. Apparently, in the following examples there were not sufficient practical or experimental grounds to reject the suspected outliers and hence some statistical judgement became necessary either to support retaining the "outliers" in the sample or leave little doubt that certain of the observations should be questioned.

EXAMPLE 1. Our first example has almost become a classical one as Irwin [3], Rider [2], and other writers on the subject including Chauvenet, Peirce, Gould, etc. (see Rider's survey [2]) all refer to it, applying their various tests. The example consists of a sample of 15 observations of the vertical semi-diameters of Venus made by Lieut. Herndon in 1846 and is given in William Chauvenet's, *A Manual of Spherical and Practical Astronomy*, II (5th ed., 1876), p 562. The individual residuals or deviations from the mean are:

−0.30″	0.48	0.63	−0.22	0.18
−0.44	−0.24	−0.13	−0.05	0.39
1.01	0.06	−1.40	0.20	0.10

Arranging the observations in increasing order of magnitude, we have:

−1.40″	−0.24	−0.05	0.18	0.48
−0.44	−0.22	0.06	0.20	0.63
−0.30	−0.13	0.10	0.39	1.01

and it is seen that two of the residuals, -1.40 and 1.01, appear to be outliers. Rider [2] indicates that the above observations have been referred to by previous writers as "residuals"; nevertheless their sum is 0.27, so that the sample mean, $\bar{x} = .018$. Let us apply the exact test, i.e. T_1 of Pearson and Chandra Sekar or S_1^2/S^2 as developed in Section 8 for a single outlier to the least observation, -1.40. We find $x_1 = -1.40$, $\bar{x} = .018$ and $s = .532$ (alternatively, we find $S^2 = 4.2496$ using all 15 observations and $S_1^2 = 2.0953$ which is based on 14 observations, the suspected outlier -1.40 not being included). Further, $T_1 = \frac{\bar{x} - x_1}{s} = \frac{.018 + 1.40}{.532} = 2.665$ (or $S_1^2/S^2 = 0.4931$) and from Table IA (or Table I) we see that $0.01 \leq P \leq 0.025$ so that we would reject the observation -1.40 when using the 5% level of significance. Having rejected -1.40, we now have left a sample of 14 observations and test the greatest one, i.e. 1.01. For T_n based on the remaining 14 observations, we have $n = 14$, $x_n = 1.01$, $\bar{x} = .119$ and $s = .387$ (alternatively, for the new sums of squares, we find $S_n^2 = 1.2409$ leaving out 1.01 and $S^2 = 2.0953$ including the observation 1.01). Hence, $T_n = \frac{x_n - \bar{x}}{s} = \frac{1.01 - .119}{.387} = 2.302$ (or $S_n^2/S^2 = 0.5922$) and from Table IA (or I), we find P slightly less than .10, so that we decide to retain the observation 1.01.

It would have been interesting nevertheless to see whether or not the test $S_{1,n}^2/S^2$ would have rejected simultaneously the observations -1.40 and 1.01 if percentage points for the distribution of this statistic were available.

It is of interest to remark that for this particular example Irwin [3, page 245], using the difference between the first two individuals divided by an estimate of σ, i.e. $\frac{x_2 - x_1}{\sigma}$, concluded also that -1.40 but not 1.01 should be rejected. In testing *both* of these observations, Irwin used the single biased estimate for σ,

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = = .5326 \qquad (\text{assuming } \bar{x} = 0),$$

based on all 15 observations. It is a mere coincidence, of course, that for this example Irwin's test gives the same result as the exact test T_1 or the test based on the ratio S_1^2/S^2. In this connection, Irwin rightly calls attention to the fact that in dealing with a sample of only 15 observations the standard deviation of the sample is a very unreliable estimate of the population standard deviation.

It is remarked that here we would, of course, hesitate to apply the test $\frac{\bar{x} - x_1}{\sigma}$ to the observation -1.40 as we do not have available and accurate estimate of σ from past data.

EXAMPLE 2. The following ranges (horizontal distances from gun muzzle to point of impact) were obtained in firing projectiles from a weapon at a constant angle of elevation and at the same weight of charge of propellant powder:

Distances in yards

4782	4420
4838	4803
4765	4730
4549	4833

It is desired to know whether the projectiles exhibit uniformity in ballistic behavior or if some of the ranges, such as 4549 and 4420, are not consistent with the others.

Arranging the distances or ranges in increasing order of magnitude,

4420	4782
4549	4803
4730	4833
4765	4838

we suspect the presence of two outliers, i.e. 4420 and 4549. Having no available knowledge of σ from past data for this example, an intuitively efficient test to apply would be that of Section 9, i.e $S_{1,2}^2/S^2$.

We find

$$\frac{S_{1,2}^2}{S^2} = \frac{\sum_{i=3}^{8} (x_i - \bar{x}_{1,2})^2}{\sum_{i=1}^{8} (x_i - \bar{x})^2} = .054$$

which is significant at the .01 level (Table V) and consequently we would judge the distances 4420 and 4549 yds. as being unusually low.

As a matter of interest and as a recommended temporary practical expedient for testing several "outliers", consider for example the last seven of the above ordered observations,

4549	4803
4730	4833
4765	4838
4782	

and apply the exact test, S_1^2/S^2, to the smallest observation, 4549. We find $S_1^2/S^2 = .145$ so that $.01 < P < .025$ from Table I and we should thus reject 4549 from the sample of seven. Moreover, we should now surely reject 4420 as being outlying, arriving at the same result we had for the test $S_{1,2}^2/S^2$. Thus, as a general temporary expedient in testing for "outliers" one could rank the observations, and apply the tests S_1^2/S^2 (or S_n^2/S^2) and $S_{1,2}^2/S^2$ (or $S_{1,n}^2/S^2$), thus working from the "inside" observations of the ranked sample in order to establish consistency of the observations.

12. Additional comments. Although we have used a significance level of .05 in the examples, it may be preferable from a practical viewpoint to reject outlying observations only at a lower level, such as .01 or .005.

Extensions of the ideas for testing outlying observations presented in this paper may lead to efficient sample criteria for testing the significance of various numbers of high, low, or simultaneously high and low sample values. However, the mathematical details would probably be complicated. In this connection, it is remarked nevertheless that the advent of high-speed computing devices may have considerable bearing on establishing experimentally any probability distribution. That is to say high-speed electronic computing devices could probably be programmed to generate random numbers with frequencies equal to those of the normal (or any other) distribution, to compute various functions (such as ratios in this paper) of sample values, etc., and establish frequency distributions to a desired order of accuracy.

13. Acknowledgement. The author is greatly indebted to Prof. C. C. Craig under whose most competent guidance this work was carried out. Indeed, the stimulus and encouragement received by the author as a result of Prof. Craig's interest in the problem were paramount in orienting various phases of the subject matter and in accomplishing the results given herein. In connection with the computing, debts of gratitude are owed to several members of the staff of the Ballistic Research Laboratories. It is desired to express appreciation to Col. Leslie E. Simon, Director of the Ballistic Research Laboratories, for recognizing the importance and desirability of carrying out the computations on high-speed computing devices. In this connection, appreciation is also expressed to Dr. L. S. Dederick, Chief, Computing Laboratory, Ballistic Research Laboratories. The programming of the $H_n(u)$ functions on the Electronic Numerical Integrator and Computor was done by Dr. Derrick Lehmer of the University of California, who was with the Computing Laboratory, BRL, during the latter part of World War II and Miss Ruth Lichterman. The author is particularly indebted to Dr. Franz Alt, Dr. Bernard Dimsdale, Miss R. Lichterman, Mr. John Holberton, Miss H. Marks, Mr. F. Spence, and others of the Computing Laboratory during the period the functions $H_n(\chi)$ were computed on the ENIAC. The computing of the distribution of the statistics S_n^2/S^2 and $S_{n-1,n}^2/S^2$ was done on the Bell Relay Computors under the direction of Mr. J. O. Harrison, Mrs. M. Masincup and Mr. E. Cushen. The author also desires to express appreciation to Miss Helen J. Coon for considerable computation and checking carried out on a table-model computing machine.

REFERENCES

[1] E. S. Pearson and C. Chandra Sekar, "The efficiency of statistical tools and a criterion for the rejection of outlying observations", *Biometrika*, Vol. 28 (1936), pp 308–320.

[2] P R. Rider, "Criteria for rejection of observations", *Washington University Studies*—New Series, Science and Technology—No. 8, St. Louis (1933).

[3] J. O. Irwin, "On a criterion for the rejection of outlying observations", *Biometrika*, Vol. 17 (1925), pp. 238-250.

[4] "Student", *Biometrika*, Vol. 19 (1927), pp 151-164.

[5] L. H. C. Tippett, "The extreme individuals and the range of samples taken from a normal population", *Biometrika*, Vol. 17 (1925), pp. 151-164.

[6] E S. Pearson, "A further note on the distribution of range in samples taken from a normal population", *Biometrika*, Vol. 18 (1926), pp. 173-194.

[7] *Tables for Statisticians and Biometricians, Part II*, edited by Karl Pearson, pp. CX-CXIX.

[8] E. S. Pearson, "The percentage limits for the distribution of range in samples from a normal population", *Biometrika*, Vol. 24 (1932), pp. 404-417.

[9] H. O. Hartley, "The range in random samples", *Biometrika*, Vol. 32 (1942), pp 334-348.

[10] E. S Pearson and H. O. Hartley, "The probability integral of the range in samples of n observations from a normal population", *Biometrika*, Vol. 32 (1942), pp. 301-310

[11] A. T. McKay, "The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe", *Biometrika*, Vol. 27 (1935), pp. 466-471.

[12] W R. Thompson, "On a criterion for the rejection of observations and the distribution of the ratio of the deviation to the sample standard deviation", *Annals of Math. Stat*, Vol. 6 (1935), pp. 214-219

[13] H. J Godwin, "On the distribution of the estimate of mean deviation obtained from samples from a normal population", *Biometrika*, Vol. 33 (1945), pp. 254-256.

[14] W.P A, *Tables of Probability Functions*, Vols. I and II, New York, N.Y (1942)

[15] H. O. Hartley, "Note on the calculation of the distribution of the estimate of mean deviation in normal samples", *Biometrika*, Vol. 33 (1945), pp. 257-258.

[16] "Tables of the probability integral of the mean deviation in normal samples", *Biometrika*, Vol. 33 (1945), pp. 259-265.

[17] J. Neyman and E S. Pearson, "On the problem of the most efficient tests of statistical hypotheses", *Phil. Trans. Roy. Soc* (London), Vol. 231 (1933), pp. 289-337.

[18] E S Pearson and H. O. Hartley, "Tables of the probability integral of the studentized range,", *Biometrika*, Vol. 33 (1943), pp. 89-99.

[19] K Pearson, *Tables of the Incomplete Beta-Function*, published by the Biometrika Office, University College, London.

[20] K. R Nair, "The distribution of the extreme deviate from the sample mean and its studentized form," *Biometrika*, Vol. 35 (1948), pp. 118-144.

DISTRIBUTION OF THE CIRCULAR SERIAL CORRELATION COEFFICIENT FOR RESIDUALS FROM A FITTED FOURIER SERIES[1,2]

By R. L. Anderson and T. W. Anderson[3]

North Carolina State College and Columbia University

CONTENTS

Summary. In this paper the observations are considered to be normally distributed with constant variance and means consisting of linear combinations of certain trigonometric functions. The likelihood ratio criterion for testing the independence of the observations against the alternatives of circular serial correlation of a given lag is found to be a function of the circular serial correlation coefficient for residuals from the fitted Fourier series (Section 4). The exact distribution (Section 5), the moments (Section 6), and approximate distributions

[1] Included in Cowles Commission Papers, New Series, No. 42.

[2] Presented to the meeting of the Institute of Mathematical Statistics at New York, December 30, 1947.

[3] Fellow of the John Simon Guggenheim Memorial Foundation; Research Consultant of the Cowles Commission for Research in Economics.

(Section 7) are given for the cases of greatest interest. From these results significance levels have been found (Section 3). The use of these levels is indicated (Section 2), and an example of their use is given (Section 3).

1. Introduction. Two mathematical models have been used extensively in time-series analysis. In one model the observation is the sum of a "systematic part" and a random error. The cyclical properties of this model result from the cyclical properties of the systematic part, which is usually taken to be a short Fourier series. The stochastic element is superimposed on the non-stochastic part, and the error at one time point does not affect a later observation. The other model is the stochastic difference equation or "autoregressive model." An observation is the sum of a linear function of previous observations and a random element. The cyclical properties follow from the properties of the difference equation (i e., the linear combination of observations), but are disturbed by the random disturbance that is integrated into the system. A more general model can be constructed that includes both of the two mentioned. The observation can be taken as a linear combination of past observations and Fourier terms plus a random element.

In this paper, the linear combination will be only a multiple of some preceding observation. For lag 1, the model is of the form

$$x_i - \mu_i = \rho(x_{i-1} - \mu_{i-1}) + u_i, \qquad i = 1, 2, \cdots, N, \tag{1}$$

where $x_0 \equiv x_n$ and $\mu_0 \equiv \mu_n$. In (1), the $\{x_i\}$ are the N observations; the $\{u_i\}$ are N random disturbances, each assumed normally and independently distributed with zero mean and variance σ^2; the means $\{\mu_i\}$ are linear combinations of some of the N functions of i: $\cos \frac{2\pi ig}{N}$ and $\sin \frac{2\pi ih}{N}$. For N odd, $g = 0, 1, \cdots, \frac{1}{2}(N - 1)$; $h = 1, \cdots, \frac{1}{2}(N - 1)$. For N even, $g = 0, 1, \cdots, \frac{1}{2}N$; $h = 1, \cdots, \frac{1}{2}N - 1$. Hence,

$$\mu_i = \sum_{g'} \alpha_{g'} \cos \frac{2\pi ig'}{N} + \sum_{h'} \beta_{h'} \sin \frac{2\pi ih'}{N}, \tag{2}$$

where g' and h' run over certain values of the ranges of g and h, respectively. Let K' be the number of terms in (2). Usually the constant term, α_0, is included (in this case $g = 0$ and $\cos \frac{2\pi ig}{N} = 1$). Of the N trigonometric functions available, the terms in (2) are usually chosen so that terms with certain periods are included and terms with other periods are excluded. It should be noted that (1) defines a circular model.

The sample estimates of $\alpha_{g'}$ and $\beta_{h'}$ are the usual regressions of x_i on $\cos \frac{2\pi ig'}{N}$ and $\sin \frac{2\pi ih'}{N}$, respectively. Because of the orthogonality of these trigonometric terms, the estimates are

$$
(3) \qquad
\begin{aligned}
a_{g'} &= \sum_{i=1}^{N} x_i \cos \frac{2\pi i g'}{N} \Big/ \frac{N}{2}, \qquad g \neq 0, \tfrac{1}{2}N, \\
b_{h'} &= \sum_{i=1}^{N} x_i \sin \frac{2\pi i h'}{N} \Big/ \frac{N}{2}, \\
a_0 &= \sum_{i=1}^{N} x_i \Big/ N, \\
a_{\frac{1}{2}N} &= \sum_{i=1}^{N} x_i \cos \pi i \Big/ N = \sum_{i=1}^{N} (-1)^i x_i \Big/ N.
\end{aligned}
$$

The fitted series is

$$
(4) \qquad m_i = \sum_{g'} a_{g'} \cos \frac{2\pi i g'}{N} + \sum_{h'} b_{h'} \sin \frac{2\pi i g'}{N}.
$$

where the sums on g' and h' are over the ranges in (2).

The serial correlation coefficient suitable for this model is

$$
(5) \qquad R = \frac{\sum_{i=1}^{N} (x_i - m_i)(x_{i-1} - m_{i-1})}{\sum_{i=1}^{N} (x_i - m_i)^2},
$$

where $m_0 \equiv m_n$. This statistic can be used to estimate ρ, or it can be used to test hypotheses about ρ. In fact, for the circular model this statistic leads to the best tests [3].

It is hoped that the mathematical model studied in this paper can be used in the treatment of certain problems in economic time series. For example, the seasonal variation in a series of data may be considered as a "systematic part" made up of trigonometric components. In the next section we discuss in a more detailed way how the use of this model may arise in the field of economics.

We have considered circular serial correlation, although in most statistical problems it is non-circular serial correlation that is involved. The reason for treating the circular case is the inherent mathematical simplicity. The circular coefficient and Fourier series of the type (2) are "naturally" related. The relevant fact is that the vectors

$$
\left(\cos \frac{2\pi g}{N}, \cos \frac{4\pi g}{N}, \cdots, \cos \frac{2N\pi g}{N}\right) \text{ and } \left(\sin \frac{2\pi h}{N}, \sin \frac{4\pi h}{N}, \cdots, \sin \frac{2N\pi h}{N}\right)
$$

are characteristic vectors of the matrix of the quadratic form in $(x_i - m_i)$ of the numerator of R. For this reason the distribution and significance points are easily obtained.

In the usual applications the circular coefficient can be used even if the hypothesis alternative to independence of observations is non-circular serial correla-

tion. The circular coefficient may not have as good power against non-circular alternatives as non-circular coefficients, such as

$$\frac{\sum_{i=2}^{N} (x_i - m_i)(x_{i-1} - m_{i-1})}{\sum_{i=1}^{N} (x_i - m_i)^2}. \tag{6}$$

However, the difference between these two statistics is $(x_1 - m_1)(x_N - m_N)/\Sigma(x_i - m_i)^2$, and it can be shown that this converges stochastically to zero (as N increases and ρ remains fixed).

2. The use of fitted Fourier series. A linear combination of trigonometric terms may be used as a regression function when there is a "systematic part" (or "trend") that is periodic. For instance, it may be reasonable to assume that a series of agricultural data has a systematic component with certain periodicities due to variation in weather. Then one may ask whether this regression function "explains" all of the interrelations in the series.

An example taken from agricultural economics is a development of that given by Koopmans [8]. Suppose p_t and q_t are the price and supply, respectively, of a given farm product at time t. Let $Q_t^{(d)}$ be the quantity demanded at time t if $p_t = P$, and $Q_t^{(s)}$ be the quantity supplied at time t if $p_{t-L} = P$, where P is an arbitrarily selected point of reference on the price scale, serving to define the Q's. Let the market equations be defined as follows:

$$p_t - P = -\beta(q_t - Q_t^{(d)}) + u_t, \tag{7}$$

$$q_t - Q_t^{(s)} = \delta(p_{t-L} - P) + v_t, \tag{8}$$

where u and v are random disturbances. The first equation expresses the price depressing tendency of an abnormally large supply; the second expresses the supply-stimulating influence of abnormally high prices L time units earlier (the time between planning the product and selling it). We can substitute from (7) at time $(t - L)$ into (8) and obtain

$$q_t - Q_t^{(s)} = \rho(q_{t-L} - Q_{t-L}^{(d)}) + w_t, \tag{9}$$

which is of the form (1) for general lag L ($i - 1$ is replaced by $t - L$) if $Q_t^{(s)} - \rho Q_{t-L}^{(d)} = \mu_t - \rho\mu_{t-L}$. Now we may wish to test the null hypothesis, $H_0 : \rho = 0$. If we assume that our alternative hypothesis is $H_a : \rho > 0$, we can test the null hypothesis by use of the positive tail of the distribution of R. Similarly for $H_a : \rho < 0$, we would use the negative tail of the distribution of R. In other cases, if we believe $\rho \neq 0$, we might wish to estimate ρ.

It is of particular interest to consider using the Fourier series for seasonal variation. The most important cases are given below with indications of the appropriate tables of significance points for testing the hypothesis $\rho = 0$. (a) *Annual data*. Here only a constant is fitted; this is the sample mean. The tables

given in [2] or [5] are to be used. (b) *Semi-annual data.* To "correct" for variation of period two we fit a constant and $\cos \pi t = (-1)^t$. The table given in Section 3 for $P = 2$ is to be used. (c) *Quarterly data.* The four terms to be fitted are 1, $\cos \pi t = (-1)^t$, $\cos \frac{\pi t}{2}$, and $\sin \frac{\pi t}{2}$. The table given in Section 3 for $P = 2$ and 4 is to be used. (d) *Bimonthly data.* The six terms to be fitted are 1, $\cos \pi t$, $\cos \frac{2\pi t}{3}$, $\sin \frac{2\pi t}{3}$, $\cos \frac{\pi t}{3}$, and $\sin \frac{\pi t}{3}$. The table given in Section 3 for $P = 2$, 3, and 6 is to be used. (e) *Monthly data.* The twelve terms to be fitted are 1, $\cos \frac{\pi t}{6}$, $\sin \frac{\pi t}{6}$, $\cos \frac{\pi t}{3}$, $\sin \frac{\pi t}{3}$, $\cos \frac{\pi t}{2}$, $\sin \frac{\pi t}{2}$, $\cos \frac{2\pi t}{3}$, $\sin \frac{2\pi t}{3}$, $\cos \frac{5\pi t}{6}$, $\sin \frac{5\pi t}{6}$, and $\cos \pi t = (-1)^t$. The table given in Section 3 for $P = 2$, 12/5, 3, 4, 6, and 12 is to be used. It is assumed here that the data are given for each time interval in a certain number of years. Then the residuals are the same as the residuals taken from means for each month or season. That is, if the data are monthly, one may compute the sample means for January, February, etc., and residuals are to be taken from the corresponding monthly means. The fitted Fourier coefficients are certain linear functions of these means.

3. Tables of significance points of R.

3.1. *Significance points of R using a seasonal trend for annual, semi-annual, bimonthly, and monthly data.* The calculations of significance points of R (lag 1 only) have been subdivided according to the number of terms included in the estimating equations, m_t. The significance points for only a constant in m_t have been tabulated in [2] and [5]. Since the main use for m_t equations involving sine and cosine terms seems to be for semi-annual, quarterly, bimonthly, and monthly data, for which N is even, the results presented in this paper are for N even. Then we will have all of the sine and cosine terms in pairs except for $\cos \pi i = (-1)^i$ and the constant term. We shall find it convenient to refer to the period $P_{g'} = N/g'$ or $P_{h'} = N/h'$ of the terms in (2).

We have calculated significance points R' exact to 3 decimal places, for $Pr\{R > R'\} = \alpha = .01, .05, .95$, and .99. The values of R' corresponding to $\alpha = .01$ and .05 are usually indicated as the positive significance points and those corresponding to $\alpha = .95$ and .99, the negative significance points. In all of these cases, except for annual data, the distribution of R is symmetrical. Hence only the positive significance points need be given, since the negative points are simply the corresponding positive points with opposite sign; that is, $R'(.95) = -R'(.05)$, $R'(.99) = -R'(.01)$.

The significance points were calculated from the exact distribution of R given in Section 5 for all N up to the values where the approximate significance points using an Incomplete Beta distribution (Section 7) were the same as the exact significance points. The Incomplete Beta significance points were used

up to the value of N for which a normal approximation was satisfactory. For some of the results, the normal points became sufficiently accurate to be used following the exact points.

The values of R' are given in Table 1 except for (a), for the following values of N:

(a) *Annual data*—see the tables in [2] or [5].

(b) *Semi-annual data* ($P = 2$): $N = 6(2)60$. The exact points were needed for N through 10($\alpha = .05$) and N through 22 ($\alpha = .01$). The normal points could be used for $N = 60$ ($\alpha = .05$) but were still too large by .003 for $N = 60$ ($\alpha = .01$)

(c) *Quarterly data* ($P = 2, 4$): $N = 8(4)100$. The exact points were needed for N through 20 ($\alpha = .05$) and N through 32 ($\alpha = .01$). The normal points were adequate for all N above 20 ($\alpha = .05$) but were still too large by .001 for $N = 100$ ($\alpha = .01$).

(d) *Bimonthly data* ($P = 2, 3, 6$): $N = 12(6)150$. The exact points were needed for N through 24 ($\alpha = .05$) and N through 30 ($\alpha = .01$). Again the normal points were adequate for all N above 24 ($\alpha = .05$) but were still too large by .0005 for $N = 150$ ($\alpha = .01$).

(e) *Monthly data* ($P = 2, 12/5, 3, 4, 6, 12$): $N = 24(12)300$. The exact points were needed for $N = 24$ ($\alpha = .05$) and $N = 24, 36$ ($\alpha = .01$) The normal points were adequate for $N > 24$ ($\alpha = .05$) and $N > 300$ ($\alpha = .01$).[4]

Significance points for the Incomplete Beta approximation (See Section 7) are tabulated in terms of $2p$ and $2q$. The values of $2p$ and $2q$ are the same when $\mu_1'(R) = 0$, for (c), (d), and (e) above these values are simply $N - 3$, $N - 5$, and $N - 11$, respectively. Hence, for two-tailed significance points for these cases, the ordinary correlation tables can be used with $N - 3$, $N - 5$, and $N - 11$ degrees of freedom, respectively. Also, our one-tailed significance points can be approximated by use of the 10% and 2% significance points for the ordinary correlation coefficient. 10%, 5%, 2%, 1%, and 0.1% two-tailed significance points have been tabulated by Fisher and Yates [6]. These significance points are accurate to three decimal places for the serial correlation coefficients as follows.[5]

(c) $n = N - 3$ degrees of freedom: $N \geq 24$ ($\alpha = .05$); $N \geq 36$ ($\alpha = .01$),

(d) $n = N - 5$ degrees of freedom: $N \geq 24$ ($\alpha = .05$); $N \geq 30$ ($\alpha = .01$),

(e) $n = N - 11$ degrees of freedom: $N \geq 24$ ($\alpha = .05$ and $\alpha = .01$), where α is the one-tailed significance point. For semi-annual data (b), $2p = 2q = \dfrac{N^2 - 3N + 4}{N - 4}$, which is not an integer for $N > 12$. When $N = 12$, $2p = 2q = 14$, for which the ordinary correlation significance point is adequate for $\alpha = .05$.

---

[4] It should be noted that for (c), (d), and (e), an approximation given by Cochran [4] is easily computed and is more accurate than the normal approximation for the $\alpha = .01$ significance points

[5] In [6] n is 2 less than the number of pairs used in computing the ordinary correlation coefficient when the sample means are first subtracted

Details of computing techniques using the exact distribution are given by R. L. Anderson [1] for computing values of R' when $\mu_s = 0$.

3.2. *Significance points of R for other single-period trends.* Significance points have also been obtained for $P = 3$, $P = 4$, $P = 6$, and $P = 12$, for which $K' = 3$.

TABLE 1

*Exact significance points, R', for different fitted series**

$P = 2$			$P = 2, 4$			$P = 2, 3, 6$			$P = 2, 12/5, 3, 4, 6, 12$		
$N\diagdown\alpha$	05	01	$N\diagdown\alpha$	.05	.01	$N\diagdown\alpha$	.05	01	$N\diagdown\alpha$	.05	01
6	.495	.499	8	.636	.693	12	.592	.744	24	.441	.592
8	.484	.607	12	.515	.661	18	.442	.592	36	.323	.445
10	.453	.601	16	.439	.582	24	.369	.504	48	.267	.371
12	.426	.572	20	.388	.523	30	.323	.445	60	.233	.325
14	.402	.544	24	.351	.478	36	.291	.403	72	.209	.293
16	.382	.519	28	.323	.441	42	.267	.371	84	.191	.268
18	.364	.496	32	.300	.414	48	.248	.346	96	.177	.249
20	.348	.476	36	.282	.391	54	.233	.325	108	.166	.234
22	.334	.458	40	.267	.371	60	.220	.308	120	.157	.221
24	.321	.442	44	.254	.354	66	.209	.293	132	.149	.210
26	.310	.427	48	.243	.338	72	.200	.280	144	.142	.200
28	.300	.414	52	.233	.325	78	.191	.268	156	.136	.192
30	.290	.402	56	.224	.313	84	.184	.258	168	.131	.184
32	.282	.390	60	.216	.302	90	.177	.249	180	.126	.178
34	.274	.380	64	.209	.293	96	.172	.241	192	.122	.172
36	.266	.370	68	.202	.284	102	.166	.234	204	.118	.166
38	.260	.361	72	.197	.276	108	.161	.227	216	.115	.162
40	.254	.353	76	191	.268	114	.157	.221	228	.111	.157
42	.248	.345	80	.186	.261	120	.153	.215	240	.108	.153
44	.242	.338	84	.182	.255	126	.149	210	252	.105	.149
46	.237	.331	88	.177	.249	132	.145	.205	264	.103	.146
48	.233	.324	92	.173	.243	138	.142	.200	276	.101	.142
50	.228	318	96	.170	.238	144	.139	.196	288	.099	.140
52	.224	.313	100	.166	.234	150	.136	.192	300	.097	.136
54	.220	.307									
56	.216	.302									
58	.212	.297									
60	.209	.292									

* P = Periods Used in Fitted Series.

In these cases, the distribution of R is asymmetrical. The Incomplete Beta approximation is symmetrical for $P = 3$, with $2p = 2q = N - 2$, even though the exact distribution is not.

The significance points for these single-period trends are given in Table 2.

The exact distribution was required to compute the $\alpha = .01$ and $.99$ significance points for N through 48 in all cases and also for most cases with $\alpha = .05$ and $.95$. For $N > 48$, the Cochran approximation [4] gave the same results as the Incomplete Beta approximation. Since this Cochran approximation can be computed more rapidly, it should be used if other significance points are desired. The normal approximation is not recommended because it is less accurate than the Cochran approximation and requires almost as much calculation. For $\alpha = .01$ and $.99$, the significance points using the normal approximation were too large (in absolute value) by from .0005 to .001 for the last entries in Table 2. The two-

TABLE 2

Exact significance points, R', for single periods > 2

$P = 3$					$P = 6$				
N	α				N	α			
	.99	.95	.05	.01		.99	.95	.05	.01
6	−.970	−.854	.496	.500					
12	−.690	−.522	.475	.619	12	−.760	−.651	.296	.506
18	−.558	−.409	.392	.526	18	−.630	−.509	.277	.440
24	−.480	−.348	.340	.463	24	−.540	−.427	.254	.393
30	−.428	−.309	.304	.417	30	−.482	−.373	.230	.359
36	−.389	−.280	.277	.382	36	−.438	−.335	.220	.332
42	−.360	−.257	.256	.356	42	−.403	−.306	.207	.311
48	−.336	−.240	.240	.334	48	−.375	−.283	.197	.294
54	−.316	−.226	.226	.316	54	−.352	−.264	.188	.279
60	−.300	−.214	.214	.300	60	−.333	−.248	.180	.266
66	−.286	−.204	.204	.286	66	−.316	−.235	.173	.255
72	−.274	−.195	.195	.274	72	−.301	−.224	.167	.246
78	−.263	−.187	.187	.263	78	−.288	−.214	.161	.237
84	−.254	−.181	.181	.254	84	−.277	−.205	.156	.229
90	−.245	−.175	.175	.245	90	−.267	−.197	.151	.222
96	−.237	−.169	.169	.237	96	−.258	−.190	.147	.216
102	−.230	−.164	.164	.230	102	−.250	−.184	.143	.210
108	−.224	−.159	.159	.224	108	−.242	−.178	.140	.205
114	−.218	−.155	.155	.218	114	−.235	−.173	.137	.200
120	−.212	−.151	.151	.212	120	−.229	−.168	.134	.195
126	−.207	−.147	.147	.207	126	−.223	−.163	.131	.191
132	−.202	−.144	.144	.202	132	−.218	−.159	.128	.187
138	−.198	−.141	.141	.198	138	−.213	−.155	.125	.183
144	−.194	−.138	.138	.194	144	−.208	−.152	.123	.180
150	−.190	−.135	.135	.190	150	−.203	−.148	.121	.177

TABLE 2—*Continued*

	$P = 4$					$P = 12$			
	α					α			
N	.99	.95	.05	.01	N	.99	.95	.05	.01
8	−.889	−.768	.503	.637	12	−.778	−.671	.096	.245
12	−.742	−.608	.420	.585	24	−.555	−.444	.197	.330
16	−.643	−.502	.369	.522	36	−.447	−.348	.188	.298
20	−.576	−.441	.333	.474	48	−.383	−.293	.175	.270
24	−.519	−.396	.306	.437	60	−.339	−.257	.163	.249
28	−.477	−.361	.285	.407	72	−.307	−.231	.153	.231
32	−.445	−.334	.268	.383	84	−.283	−.212	.145	.217
36	−.418	−.312	.253	.363	96	−.263	−.196	.138	.206
40	−.395	−.293	.241	.345	108	−.247	−.183	.132	.196
44	−.375	−.277	.230	.330	120	−.233	−.173	.126	.187
48	−.358	−.264	.221	.317	132	−.221	−.164	.121	.180
52	−.343	−.252	.213	.305	144	−.211	−.156	.117	.173
56	−.330	−.242	.206	.294	156	−.202	−.149	.113	.167
60	−.319	−.233	.199	.285	168	−.194	−.143	.110	.162
64	−.308	−.225	.193	.277	180	−.187	−.138	.107	.157
68	−.298	−.218	.188	.269	192	−.181	−.133	.104	.153
72	−.289	−.211	.183	.262	204	−.175	−.128	.101	.149
76	−.281	−.205	.178	.255	216	−.170	−.124	.099	.145
80	−.274	−.199	.174	.249	228	−.165	−.121	.097	.141
84	−.267	−.194	.170	.243	240	−.161	−.117	.094	.138
88	−.261	−.189	.166	.238	252	−.157	−.114	.092	.135
92	−.255	−.184	.162	.233	264	−.153	−.111	.091	.132
96	−.249	−.180	.159	.228	276	−.149	−.109	.089	.130
100	−.244	−.176	.156	.223	288	−.146	−.106	.087	.127
108	−.234	−.169	.150	.215	300	−.143	−.104	.086	.125
120	−.221	−.160	.143	.205					
132	−.210	−.152	.136	.196					
144	−.201	−.145	.131	.187					

tailed significance points cannot be obtained from the ordinary correlation tables except for $P = 3$.

3.3. *Example of use of significance points.* As an example of the use of these significance points, R', we shall consider the following data [17] on the receipts of butter (in units of 1,000,000 pounds) at five markets (Boston, Chicago, San Francisco, Milwaukee, and St. Louis). The figures in parentheses are deviations from the average of the given months over the 3 years.

Month	Year 1935	1936	1937	Total T_i	Average
Jan.	48.9(2.4)	48.3(1.8)	42.4(−4.1)	139.6	46.5
Feb.	43.4(−0.6)	47.1(3.1)	41.4(−2.6)	131.9	44.0
March	43.8(−4.6)	52.4(4.0)	49.0(0.6)	145.2	48.4
April	50.8(−1.5)	55.3(3.0)	50.8(−1.5)	156.9	52.3
May	67.6(1.6)	64.7(−1.3)	65.8(−0.2)	198.1	66.0
June	83.7(0.7)	79.5(−3.5)	85.9(2.9)	249.1	83.0
July	82.7(10 7)	62.6(−9.4)	70.6(−1.4)	215.9	72.0
Aug.	60.8(4.8)	51.3(−4.7)	55.8(−0.2)	167.9	56.0
Sept.	55.4(3.6)	51.0(−0.8)	49.1(−2.7)	155.5	51.8
Oct.	48.4(−1.0)	54.0(4.6)	45.7(−3.7)	148.1	49.4
Nov.	37.7(−4.5)	45.2(3.0)	43.8(1.6)	126.7	42.2
Dec.	41.0(−3.2)	44.9(0.7)	46.7(2.5)	132.6	44.2
Total	664.2(8.4)	656.3(0.5)	647.0(−8.8)	1967.5	655.8
Average	55 35(0.70)	54.69(0.04)	53.92 (−0.73)	163.96	54.65

We assume that the trend is composed of the 12 terms having periods that divide 12. We shall test the null hypothesis that the deviations from the trend are independently distributed against the alternative that there is positive serial correlation. The fitted series is of the form

$$m_i = b_0^* + \sum_{j=1}^{5} \left(b_{2j-1}^* \cos \frac{\pi i j}{6} + b_{2j}^* \sin \frac{\pi i j}{6} \right) + b_{11}^* \cos \pi i; \tag{10}$$

here we find it convenient to use the notation, $b_0^*, b_1^*, \cdots, b_{11}^*$, for the coefficients (with a different relationship between the subscripts and the trigonometric functions than in (4)). We find that the m_i are simply the average receipts given for each month in the above table (46.5, 44.0, $\cdots$, 44.2). Hence the deviations $(x_i - m_i)$ are given by the figures in parentheses (2.4, −0.6, $\cdots$, 2.5). The calculated lag 1 circular serial correlation coefficient is

$$\begin{aligned} R_0 &= \frac{(2.4)(-0.6) + (-0.6)(-4.6) + \cdots + (1.6)(2.5) + (2.5)(2.4)}{(2.4)^2 + (-0.6)^2 + \cdots + (2.5)^2} \\ &= \frac{232.18}{474.51} = 0.489. \end{aligned} \tag{11}$$

Entering Table 1 for $P = 2, 12/5, 3, 4, 6$, and 12 and $N = 36$, we find that $R'(.05) = 0.323$ and $R'(.01) = .445$. Hence, at either the 5% or 1% level the null hypothesis of zero serial correlation ($\rho = 0$) is to be rejected (against the alternative single-tail hypothesis, $\rho > 0$). If we had been interested in the two-

tailed alternative hypothesis, $\rho \neq 0$, we would use the ordinary correlation tables with $N - 11 = 25$ degrees of freedom and we would find that for the two-tailed test $R'(.01) = 0.487$. Our value is significant at the 5% level and barely significant at the 1% level.

The values of h^* in (10) are computed as follows

$$
\begin{aligned}
h_0^* &= \sum_{i=1}^{12} T_i/36, \\
h_{2j-1}^* &= \sum_{i=1}^{12} T_i \cos \frac{\pi i j}{6} \Big/ 18, \\
h_{2j}^* &= \sum_{i=1}^{12} T_i \sin \frac{\pi i j}{6} \Big/ 18, \\
h_{11}^* &= \sum_{i=1}^{12} T_i \cos \pi i/36
\end{aligned} \tag{12}
$$

The computed values of h_0^* to h_{11}^* are 54.65, −14.82, −2.02, 6.60, 1.23, −3.98, 0.30, 2.21, 1.73, [illegible]0.61, [illegible], 0.15, respectively. However, it is not necessary to compute these values in order to obtain m_1. The problem of estimating the variances of these h's will be discussed in Section 4.

4. Testing the hypothesis of lack of serial correlation.

4.1. *Statement of the problem.* Consider the N random variables $u_1, \cdots, u_N$, each normally and independently distributed with mean 0 and variance σ^2. Define the N variables $x_1, \cdots, x_N$ by the equations

$$x_i - \mu_i = \rho(x_{i-L} - \mu_{i-L}) + u_i \qquad (i = 1, \cdots, N), \tag{13}$$

where

$$x_{-j} = x_{N-j}, \quad \mu_{-j} = \mu_{N-j} \qquad (j = 0, 1, \cdots, N-1) \tag{14}$$

and μ_i is the linear combination of trigonometric functions given in (2). If L and N are relatively prime (in particular, if $L = 1$), the Jacobian of the transformation from $\{u_i\}$ to $\{x_i\}$ is $1 - \rho^N$, and the probability density of $\{x_i\}$ is

$$\frac{1 - \rho^N}{(2\pi\sigma^2)^{\frac{1}{2}N}} e^{-\frac{1}{2}Q/\sigma^2}, \tag{15}$$

where $Q = (1 + \rho^2) \sum_{i=1}^{N} (x_i - \mu_i)^2 - 2\rho \sum_{i=1}^{N} (x_i - \mu_i)(x_{i-L} - \mu_{i-L})$. If $L = 1$, the covariance between x_i and x_j is $\sigma^2[\rho^{|i-j|} + \rho^{N-|i-j|}]/[(1 - \rho^N)(1 - \rho^2)]$. If $L = q\alpha$ and $N = p\alpha$, where p, q, and α are positive integers and q and p are relatively prime, then the Jacobian is $(1 - \rho^p)^\alpha$ and the density of $\{x_i\}$ is

$$\frac{(1 - \rho^p)^\alpha}{(2\pi\sigma^2)^{\frac{1}{2}N}} e^{-\frac{1}{2}Q/\sigma^2}. \tag{16}$$

We shall now obtain the likelihood ratio test of the hypothesis $H_0 : \rho = 0$ on the basis of a sample consisting of one observation on each x_i.

4.2. *Preliminary transformations.* We shall find it convenient to express μ_i in terms of fixed variates ϕ_{ij}, having certain properties. Later we will verify that the ϕ's are simply constant multiples of the trigonometric terms in (2). We suppose now that

$$\mu_i = \sum_{j=1}^{K'} \phi_{ij} \gamma_j \qquad (i = 1, \cdots, N), \tag{17}$$

where $K' < N$, the $\{\gamma_j\}$ are parameters, and the ϕ_{ij} are known functions of i and j satisfying

$$\phi_{i-L,j} + \phi_{i+L,j} = 2\lambda_{Lj}\phi_{ij} \qquad (i = 1, \cdots, N; \quad j = 1, \cdots, K'), \tag{18}$$

$$\sum_{i=1}^{N} \phi_{ij}\phi_{ik} = \delta_{jk} \qquad (j, k = 1, \cdots, K'), \tag{19}$$

$$\phi_{-i,j} = \phi_{N-i,j} \qquad (i = 0, 1, \cdots, N - 1), \tag{20}$$

and δ_{jk} is the Kronecker delta. Let

$$m_i = \sum_{j=1}^{K'} \phi_{ij} c_j, \tag{21}$$

where

$$c_j = \sum_{i=1}^{N} x_i \phi_{ij}. \tag{22}$$

Then by usual regression theory we have

$$\sum_{i=1}^{N} (x_i - m_i)\phi_{ij} = 0, \tag{23}$$

$$\sum_{i=1}^{N} (x_i - \mu_i)^2 = \sum_{i=1}^{N} (x_i - m_i)^2 + \sum_{j=1}^{K'} (c_j - \gamma_j)^2 \tag{24}$$

because c_j is the least squares estimate of γ_j. Let us evaluate

$$\begin{aligned}
{}_L\bar{C} &= \sum_{i=1}^{N} (x_i - \mu_i)(x_{i-L} - \mu_{i-L}) \\
&= \sum_{i=1}^{N} [(x_i - m_i) + (m_i - \mu_i)][(x_{i-L} - m_{i-L}) + (m_{i-L} - \mu_{i-L})] \\
&= \sum_{i=1}^{N} (x_i - m_i)(x_{i-L} - m_{i-L}) + \sum_{i=1}^{N} \sum_{j=1}^{K'} \phi_{i-L,j}(c_j - \gamma_j)(x_i - m_i) \\
&\quad + \sum_{i=1}^{N} \sum_{j=1}^{K'} \phi_{ij}(c_j - \gamma_j)(x_{i-L} - m_{i-L}) \\
&\quad + \sum_{i=1}^{N} \sum_{j,k=1}^{K} \phi_{ik}\phi_{i-L,j}(c_k - \gamma_k)(c_j - \gamma_j).
\end{aligned} \tag{25}$$

Call the first term on the right hand side of (25) $_LC$. In view of (20) the next two terms are

$$\sum_{j=1}^{K'} \sum_{i=1}^{N} (x_i - m_i)(\phi_{i-L,j} + \phi_{i+L,j})(c_j - \gamma_j). \tag{26}$$

This is seen to be zero by consideration of (18) and (23). The last term can be written

$$\frac{1}{2} \sum_{i=1}^{N} \sum_{j,k=1}^{K'} (\phi_{ik}\phi_{i-L,j} + \phi_{i+L,j}\phi_{ik})(c_k - \gamma_k)(c_j - \gamma_j) = \sum_{j=1}^{K'} \lambda_{Lj}(c_j - \gamma_j)^2 \tag{27}$$

by use of (18), (19), and (20). Thus

$$_LC = \sum_{i=1}^{N} (x_i - m_i)(x_{i-L} - m_{i-L}) + \sum_{j=1}^{K'} \lambda_{Lj}(c_j - \gamma_j)^2. \tag{28}$$

It follows that

$$\begin{aligned} Q = (1 + \rho^2) \sum_{i=1}^{N} (x_i - m_i)^2 - 2\rho \sum_{i=1}^{N} (x_i - m_i)(x_{i-L} - m_{i-L}) \\ + \sum_{j=1}^{K'} (1 + \rho^2 - 2\rho\lambda_{Lj})(c_j - \gamma_j)^2. \end{aligned} \tag{29}$$

We can complete the matrix $\Phi = (\phi_{ij})$ so that Φ is an N-th order square matrix with elements satisfying (18), (19), and (20). If we make the transformation

$$x_i = \sum_{j=1}^{N} \phi_{ij} c_j \qquad (i = 1, \cdots, N), \tag{30}$$

then

$$\sum_{i=1}^{N} (x_i - m_i)^2 = \sum_{j=K'+1}^{N} c_j^2, \tag{31}$$

$$\sum_{i=1}^{N} (x_i - m_i)(x_{i-L} - m_{i-L}) = \sum_{j=K'+1}^{N} \lambda_{Lj} c_j^2. \tag{32}$$

4.3. *The likelihood ratio criterion*. To obtain the likelihood ratio test of the hypothesis $H_0 : \rho = 0$ against alternative hypotheses $H_a : \rho \neq 0$, we divide the maximum of the likelihood assuming H_0 by the maximum of the likelihood assuming H_a. It is clear from (15) and (29) that if H_0 is true, the maximum likelihood estimates of γ_j and σ^2 are c_j and

$$s_0^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - m_i)^2, \tag{33}$$

respectively. If H_a is true, the maximum likelihood estimate of γ_j is c_j. To state the maximum likelihood estimates of σ^2 and ρ under H_a it is convenient to define $_LR$, the sample serial coefficient of lag L, as

$$_LR = \frac{1}{Ns_0^2} \sum_{i=1}^{N} (x_i - m_i)(x_{i-L} - m_{i-L}). \tag{34}$$

Then the maximum likelihood estimate of σ^2 under H_0 is

$$s^2 = s_0^2(1 + \hat{\rho}^2 - 2\hat{\rho}_L R), \tag{35}$$

where $\hat{\rho}$ is the maximum likelihood estimate of ρ and satisfies

$$_L R(1 + \hat{\rho}^N) - \hat{\rho}(1 + \hat{\rho}^{N-2}) = 0, \tag{36}$$

if L and N are relatively prime and satisfies

$$_L R(1 + \hat{\rho}^p) - \hat{\rho}(1 + \hat{\rho}^{p-2}) = 0, \tag{37}$$

if $L = q\alpha$, $N = p\alpha$, and p and q are relatively prime.

Upon substituting these estimates into the likelihood function we find that the likelihood ratio criterion is

$$\lambda = \frac{(1 + \hat{\rho}^2 - 2\hat{\rho}_L R)^{\frac{1}{2}N}}{1 - \hat{\rho}^N}, \tag{38}$$

if L and N are relatively prime and

$$\lambda = \left[\frac{(1 + \hat{\rho}^2 - 2\hat{\rho}_L R)^{\frac{1}{2}p}}{1 - \hat{\rho}^p}\right]^\alpha, \tag{39}$$

if $L = q\alpha$, $N = p\alpha$ and p and q are relatively prime. The maximum likelihood estimate of ρ is the root of (36) or (37) that makes (38) or (39), respectively, a minimum. It should be noticed that throughout this section ρ could be replaced by $1/\rho$ (and changing σ^2 by a factor $1 + \rho^2$). To make the maximum likelihood estimate unique, we require that $|\hat{\rho}| \leq 1$. It can be shown that there exists one and only one root of (36) or (37) that satisfies this requirement and minimizes λ. (There is a peculiarity to this solution in that if N is odd, $L = 1$, and $_1R < -1 + 2/N$, then $\hat{\rho} = -1$ is the root minimizing λ). In any case, λ is a function of $_LR$. We have shown that for $0 < {_LR} < 1$, it is a monotonic decreasing function; and for $-1 < {_LR} < 0$, it is a monotonic increasing function. A critical region defined by $\lambda \leq \lambda_0$ can, therefore, be defined by $_LR \leq R_1 < 0$ and $0 < R_2 \leq {_LR}$. (The probability that $_LR = -1$ or $+1$ is 0.) Thus we can use $_LR$ to test the null hypothesis $H_0 : \rho = 0$ instead of the likelihood ratio criterion (against one-sided alternatives they are equivalent). The strongest justification for the use of $_LR$ in testing $H_0 : \rho = 0$ is that for circular distributions the uniformly most powerful tests against one-sided alternatives and the B_1 test against two-sided alternatives are given in terms of inequalities on $_LR$ [3].

We can also use $_LR$ as an estimate of ρ. In fact, $_LR$ is asymptotically a root of (36) or (37) This is proved by showing that $_LR(1 + {_LR^N}) - {_LR}(1 + {_LR^{N-2}}) = {_LR^{N-1}}(1 - {_LR^2})$ converges stochastically to zero. We shall use $_LR$ both to estimate ρ and to test hypotheses about this parameter.[6]

Now we shall define ϕ_{ij} used in Section 4.2 in terms of the trigonometric terms indicated in Section 1. In the rest of the paper we shall let the index g run from

[6] W. J. Dixon [5] arrived at $_LR$ as the maximum likelihood estimate for μ_i a constant by neglecting the Jacobian in (15).

0 to $\frac{1}{2}N$ for N even and from 0 to $\frac{1}{2}(N-1)$ for N odd; we let the index h run from 1 to $\frac{1}{2}N - 1$ for N even and from 1 to $\frac{1}{2}(N-1)$ for N odd. We shall use a prime to denote an index running over those values corresponding to fitted terms and a double prime to denote an index running over those values corresponding to terms not fitted.

Let the N trigonometric functions of i, namely $\cos \frac{2\pi ig}{N}$ and $\sin \frac{2\pi ih}{N}$, be numbered from 1 to N such that the fitted terms are numbered from 1 to K' and the non-fitted terms from $K' + 1$ to N. According to this numbering we define ϕ_{i1} to ϕ_{iN} as

$$\phi_{ij} = \sqrt{\frac{2}{N}} \cos \frac{2\pi ig}{N}, \tag{40}$$

or

$$\phi_{ij} = \sqrt{\frac{2}{N}} \sin \frac{2\pi ih}{N}. \tag{41}$$

Defined this way, the ϕ_{ij} satisfy (18) and (19) and (20). It can be shown by using the addition formulas for sines and cosines that

$$\lambda_{L,j} = \cos \frac{2\pi Lf}{N}, \tag{42}$$

where $f = g$ or $f = h$ depending on whether j refers to a term (40) or (41). We shall assume that the numbering of trigonometric functions is such that

$$\lambda_{L,K'+1} \geq \lambda_{L,K'+2} \geq \cdots \geq \lambda_{L,N}. \tag{43}$$

It can easily be seen that (2) is of the form (17) except that $\alpha_{g'}$ and $\beta_{h'}$ must be multiplied by $\sqrt{\frac{1}{2}N}$ unless $g' = 0$ or $\frac{1}{2}N$ and by $\sqrt{N}$ for $g' = 0, \frac{1}{2}N$ to obtain γ_j. The regression coefficients $a_{g'}$ and $b_{h'}$ are similarly related to the c_j.

It can be seen from (29) that the a_f and b_f are independently distributed with variance $\frac{1}{2}N\sigma^2 \Big/ \left(1 + \rho^2 - 2\rho \cos \frac{2\pi Lf}{N}\right)$ for $f \neq 0, \frac{1}{2}N$ and variance $N\sigma^2/(1-\rho)^2$ for $f = 0$ and for $f = \frac{1}{2}N$ if L is even and $N\sigma^2/(1+\rho)^2$ for $f = \frac{1}{2}N$ if L is odd. In these variance formulas we can estimate σ^2 from (35) using ${}_LR$ for $\hat{\rho}$ and ρ.

5. The exact distribution of ${}_LR$.

5.1. *Introduction.* Under the null hypothesis $H_0 : \rho = 0$ the observations $\{x_i\}$ are normally and independently distributed with variance σ^2 and means $Ex_i = \mu_i$. The variables c_j defined by (22) and (29) are normally and independendy distributed with variance σ^2 and means γ_j. For $j > K'$, $\gamma_j = 0$. It follows from (31), (32), (33), and (34) that

$${}_LR = \frac{\sum_{j=K'+1}^{N} \lambda_{Lj} c_j^2}{\sum_{j=K'+1}^{N} c_j^2}, \tag{44}$$

where the $\lambda_{L,j}$ are given by (42) corresponding to the $K'' = (N - K')$ trigonometric terms not fitted. Thus to obtain the distribution of ${}_LR$ we need only consider the joint distribution of $\{c_j\}$, $j = K' + 1, \cdots, N$. If H_0 is true, the joint density of all the c_j is (15), where

$$Q = (1 + \rho^2)V - 2\rho\, {}_LC + \sum_{j=1}^{K'} (1 + \rho^2 - 2\rho\lambda_{L,j})(c_j - \gamma_j)^2, \tag{45}$$

and

$$V = \sum_{j=K'+1}^{N} c_j^2 \quad \text{and} \quad {}_LC = \sum_{j=K'+1}^{N} \lambda_{L,j} c_j^2.$$

5.2. *Some special distributions of* ${}_1R = R$. If the constant term ($g' = 0$) is fitted and the other terms are fitted in pairs $\left(\cos \frac{2\pi i f}{N} \text{ and } \sin \frac{2\pi i f}{N}\right)$, then K' is odd. If N is odd, then K'' is even; the λ_{1j} occur in pairs and we can define λ_k'' as

$$\begin{aligned} \lambda_{1,K'+1} = \lambda_{1,K'+2} = \lambda_1'' > \lambda_{1,K'+3} = \lambda_{1,K'+4} \\ = \lambda_2'' > \cdots > \lambda_{1,N-1} = \lambda_{1N} = \lambda_{\frac{1}{2}K''}'' . \end{aligned} \tag{46}$$

This also holds if N is even and if, in addition to the constant term and paired cosines and sines, we fit $\cos \pi i = (-1)^i (g' = N/2)$. If N is even and we do not fit $\cos \pi i$, we have K'' odd. Then

$$\begin{aligned} \lambda_{1,K'+1} = \lambda_{1,K'+2} = \lambda_1'' > \lambda_{1,K'+3} = \lambda_{1,K'+4} = \lambda_2'' > \cdots > \lambda_{1,N-2} \\ = \lambda_{1,N-1} = \lambda_{\frac{1}{2}(K''-1)}'' > \lambda_{1N} = \lambda_{\frac{1}{2}(K''+1)}'' = -1. \end{aligned} \tag{47}$$

The general expression for the distribution of R in these cases has been found by one of the authors [2]. In this case the cumulative distribution function is 1 minus

$$\begin{gathered} Pr\{R > R'\} = \sum_{k=1}^{m} (-1)^{k+1} \mid V_k \mid (\lambda_k'' - R')^{\frac{1}{2}K''-1}, \\ \lambda_{m+1}'' \leq R' \leq \lambda_m'', \end{gathered} \tag{48}$$

where V_k is found from a result of Lehmann [9] to be

$$V_k = \frac{2^{\frac{1}{2}(N+1)}}{N} \sin \frac{2\pi f''}{N} \sin \frac{\pi f''}{N} \prod_{j'} \sqrt{(\lambda_k'' - \lambda_{1j'})}, \tag{49}$$

where f'' is such that $\lambda_k'' = \cos \dfrac{2\pi f''}{N}$ and the product on j' is over the K' terms $\lambda_{1j'}$, excluding $\lambda_{1j'} = 1$. Hence, $\lambda_{1j'}$ takes on $K' - 1$ values in $\frac{1}{2}(K' - 1)$ pairs if K' is odd and in $\frac{1}{2}(K' - 2)$ pairs plus a single $\lambda_{1,j'} = -1$ if K' is even. We can also write V_k as

$$V_k = \frac{2^{\frac{1}{2}(N+K')}}{N} \sin\frac{2\pi f''}{N} \sin\frac{\pi f''}{N} \prod_{g' \neq 0} \sqrt{\sin\frac{\pi(g'+f'')}{N} \sin\frac{\pi(g'-f'')}{N}} \cdot \prod_{h'} \sqrt{\sin\frac{\pi(h'+f'')}{N} \sin\frac{\pi(h'-f'')}{N}}. \tag{50}$$

5.3. *Some special distributions of ${}_LR$ for $L > 1$.* We have noted in (44) above that $\lambda_{L,j} = \cos\frac{2\pi L f''}{N}$, where f'' corresponds to a term not used in the estimation equations for m_i, which was a function of $\left\{\cos\frac{2\pi i g'}{N}, \sin\frac{2\pi i h'}{N}\right\}$. If L, the lag, is relatively prime to N, the distribution is the same as that given above for $L = 1$, except for the re-evaluating of the λ_k''. In the article by R. L. Anderson [2], where only the constant term in m_i was used, the λ_k'' for lag L were exactly the same as the λ_k'' for lag 1. However, this will not be the case for other terms used in m_i. For example, consider lag 2 and N odd with m_i consisting of the constant term plus terms in $\cos\frac{2\pi i}{N}$ and $\sin\frac{2\pi i}{N}$. In this case the λ_k'' for lag 1 are $\left\{\cos\frac{4\pi}{N}, \cos\frac{6\pi}{N}, \cdots, \cos\frac{(N-1)\pi}{N}\right\}$ and the λ_k'' for lag 2 are

$$\left\{\cos\frac{2\pi}{N}, \cos\frac{6\pi}{N}, \cos\frac{8\pi}{N}, \cdots, \cos\frac{(N-1)\pi}{N}\right\}.$$

Next suppose the highest common factor of L and N is α (as before, $L = q\alpha$ and $N = p\alpha$, with p and q relatively prime). In this case

$$\lambda_{L,j} = \cos\frac{2\pi q f''}{p}. \tag{51}$$

minist

Since p and q are relatively prime, the results are the same as for q replaced by 1 and L replaced by α. Each root is repeated α times.

$N = 2L (p = 2)$

If we let $N = 2L$, $\lambda_k'' = \cos \pi k = +1$ or -1. $\lambda'' = +1$ corresponds to these fitted terms in m_i: $\left\{1, \cos\frac{2\pi i g'}{N}, \sin\frac{2\pi i h'}{N}\right\}$ for g', h' even. $\lambda'' = -1$ corresponds to these terms: $\left\{\cos\frac{2\pi i g'}{N}, \sin\frac{2\pi i h'}{N}\right\}$ for g', h' odd. Let $L - n_1$ be the number of terms pertaining to $\lambda'' = +1$ and $L - n_2$ be the number of terms for $\lambda'' = -1$. Then, as in [2], we have the density

$$D({}_LR_2) = \frac{(1 - {}_LR_2)^{\frac{1}{2}(n_2-2)}(1 + {}_LR_2)^{\frac{1}{2}(n_1-2)}}{2^{\frac{1}{2}(n_1+n_2)-1}\beta(\frac{1}{2}n_1, \frac{1}{2}n_2)}, \tag{52}$$

where ${}_LR_2$ was the notation used for lag L and $p = 2$. The cumulative function is the Incomplete Beta function, found by setting $x = \frac{1}{2}(1 - R')$.

$N = 3L (p = 3)$

If we let $N = 3L$, $\lambda_k'' = \cos \frac{2\pi f''}{N} = +1, -\frac{1}{2}$. The fitted terms in m_i corresponding to $\lambda'' = 1$ are $\left\{1, \cos \frac{2\pi i g'}{N}, \sin \frac{2\pi i h'}{N}\right\}$ for $g', h' = 3m$. Similarly, those corresponding to $\lambda'' = -\frac{1}{2}$ have $g', h' = 3m - 1$ or $3m - 2$. Let the number of fitted terms with $\lambda'' = +1$ be $L - n_1$ and with $\lambda'' = -\frac{1}{2}$ be $2L - n_2$. Then

$$D({}_LR_3) = \frac{(1 - {}_LR_3)^{\frac{1}{2}(n_2-2)}(\frac{1}{2} + {}_LR_3)^{\frac{1}{2}(n_1-2)}}{(3/2)^{\frac{1}{2}(n_1+n_2)-1}\beta(\frac{1}{2}n_1, \frac{1}{2}n_2)}, \tag{53}$$

where ${}_LR_3 \geq -\frac{1}{2}$. This cumulative function is also an Incomplete Beta function, found by setting $x = 2(1 - R')/3$.

$N = 4L (p = 4)$

If $N = 4L$, $\lambda_k'' = \cos \frac{2\pi f''}{N} = +1, 0, -1$. The fitted terms in m_i corresponding to $\lambda'' = 1$ have $f'' = 4m$, those for $\lambda'' = -1$ have $f'' = 4m - 2$; and those for $\lambda'' = 0$ have $f'' = 4m - 1$ or $4m - 3$. Let the number of terms in m_i of each sort be $L - n_1$, $L - n_2$, and $2L - n_3$, respectively. Then

$$D(R) = c \begin{cases} (1 + R)^{\frac{1}{2}(n_1+n_3-2)} \int_{y=0}^{1} y^{\frac{1}{2}(n_3-2)} (1 - y)^{\frac{1}{2}(n_1-2)} \\ \qquad \cdot [(1 - R) - y(1 + R)]^{\frac{1}{2}(n_2-2)}\, dy, & \text{for } R \leq 0, \\ (1 - R)^{\frac{1}{2}(n_2+n_3-2)} \int_{y=0}^{1} y^{\frac{1}{2}(n_3-2)} (1 - y)^{\frac{1}{2}(n_2-2)} \\ \qquad \cdot [(1 + R) - y(1 - R)]^{\frac{1}{2}(n_1-2)}\, dy, & \text{for } R \geq 0, \end{cases} \tag{54}$$

where R is ${}_LR_4$ and $c = \Gamma(\frac{1}{2}[n_1 + n_2 + n_3])/[\Gamma(\frac{1}{2}n_1)\Gamma(\frac{1}{2}n_2)\Gamma(\frac{1}{2}n_3)2^{\frac{1}{2}(n_1+n_2-2)}]$.

5.4. *The exact distribution of ${}_LR$ when $\rho \neq 0$.* The joint distribution of the observations for lag 1 when the null hypothesis is not true ($\rho \neq 0$) is (15), where Q is given by (45) with $L = 1$ and ${}_1C = RV$. V, R, $\{c_j\}$ $(j = 1, \cdots, K')$ are a sufficient set of statistics for estimating σ^2, ρ, and $\{\gamma_j\}$ $(j = 1, \cdots, K')$. Using the results given by Madow [11], it can be shown that the simultaneous distribution of V and R is

$$\frac{1 - \rho^N}{2^{\frac{1}{2}K''}\Gamma(\frac{1}{2}K'') \sqrt{\prod_{j'=1}^{K'} (1 + \rho^2 - 2\rho\lambda_{1j'})}} V^{\frac{1}{2}K''-1} e^{-V(1+\rho^2-2\rho R)/2\sigma^2} D(R), \tag{55}$$

where $D(R)$ is the density function corresponding to (48). Integrating V from 0 to ∞, we obtain as the density for R

$$\frac{(1 - \rho^N)(\frac{1}{2}K'' - 1)}{\sqrt{\prod_{j'=1}^{K'} (1 + \rho^2 - 2\rho\lambda_{1j'})}} (1 + \rho^2 - 2\rho R)^{\frac{1}{2}K''} \cdot \sum_{k=1}^{m} (-1)^{k+1}(\lambda_k'' - R)^{\frac{1}{2}(K''-4)} \mid V_k \mid, \tag{56}$$

for $\lambda''_{m+1} \leq R \leq \lambda''_m$, where V_k are given by (50). In the same way, one obtains the distribution of ${}_LR$ for $\rho \neq 0$ when $N = 2L$, $N = 3L$, and $N = 4L$ by multiplying (52), (53), and (54), respectively, by

$$(1 - \rho^p)^L \frac{(1 + \rho^2 - 2\rho R)^{\frac{1}{2}K''}}{\sqrt{\prod_{j'=1}^{K'} (1 + \rho^2 - 2\rho\lambda_{Lj'})}}, \tag{57}$$

where $K'' = n_1 + n_2$ or $n_1 + n_2 + n_3$. This method was used by Madow for residuals from the sample mean [12].

6. Moments.

6.1. *The exact moments of* R. Most of the results of this section are straightforward adaptations of earlier results for the case of μ_i constant. Hence, we shall omit the details of derivations. The moment generating function of V and C for $\sigma^2 = 1$ is

$$\phi(t_0, t) = E(e^{t_0 V + tC}) = \frac{1 - \rho^N}{\prod_{j''=K'+1}^{N} \left[1 + \rho^2 - 2t_0 - 2(\rho + t)\, \lambda_{1,j''}\right]^{\frac{1}{2}}}. \tag{58}$$

The h^{th} moment of $R = C/V$ is given by

$$\mu_h'(R) = \int_{-\infty}^{0} \int_{-\infty}^{y_{h-1}} \cdots \int_{-\infty}^{y_1} \frac{\partial^h \phi}{\partial t^h}\Big]_{t=0} dt_0 \prod_{i=1}^{h-1} dy_i\,, \tag{59}$$

with the $\{y_i\}$ restricted from being too large (not more than a certain amount larger than zero). In the case of independence, ($\rho = 0$), we have the following first two moments of R:

$$\begin{aligned} \mu_1'(R) &= \frac{1}{K''} \sum_{j''=K'+1}^{N} \lambda_{1,j''}; \\ \mu_2'(R) &= \frac{2}{K''\,(K''+2)} \sum_{j''=K'+1}^{N} \lambda_{1,j''}^2 + \frac{K''}{K''+2}\, [\mu_1'(R)]^2. \end{aligned} \tag{60}$$

If the $\lambda_{1,j''}$ are symmetrical (i.e. for each $\lambda_{1,j''}$, there is a $\lambda_{1,k''} = -\lambda_{1,j''}$), the mean of R is 0. For example, if 1 and $(-1)^i$ are fitted for N even, the mean is 0.

6.2. *Approximate moments of* R *when* $\rho = 0$. Since R and V are independent [8] when $\rho = 0$, $\mu'(R) = \mu'(C)/\mu'(V)$. V is a sum of squares and its moments are the same as for χ^2 with $N - K' = K''$ degrees of freedom. Using methods similar to those given by Dixon [5], we see that the moment generating function for C is

$$\phi(t) = \alpha(t) \cdot \beta(t) \cdot \gamma(t), \tag{61}$$

where

$$\begin{aligned} \alpha(t) &= \left(\frac{2}{A}\right)^{\frac{1}{2}N}, \beta(t) = A^N/[A^N - (2t)^N], \\ \gamma(t) &= \textstyle\prod_{j'} (1 - 2t\, \lambda_{1,j'}), \text{ and } A = 1 + \sqrt{1 - 4t^2}. \end{aligned} \tag{62}$$

In this case, $\lambda_{1,j'} = \cos \frac{2\pi j'}{N}$ includes all K' terms corresponding to those in m_i. Since the first N derivatives of $\beta(t)$ are zero at $t = 0$, we can use

$$\tilde{\phi}(t) = \alpha(t) \cdot \gamma(t) = \frac{2^{\frac{1}{2}N} \prod_{j'} (1 - 2t\,\lambda_{1,j'})^{\frac{1}{2}}}{(1 + \sqrt{1 - 4t^2})^{\frac{1}{2}N}} \tag{63}$$

as an approximation to (61). This expression yields the exact moments of C up to order N.

As a special case, consider $K' = 3$, with $\lambda_{1,1} = 1$ and $\lambda_{1,2} = \lambda_{1,3} = \cos \frac{2\pi j^*}{N}$. In this case

$$\tilde{\phi}_3(t) = \left(1 - 2t \cos \frac{2\pi j^*}{N}\right) \tilde{\phi}_1(t). \tag{64}$$

Successive derivatives of (64) at $t = 0$ show that

$$\mu_h'(R_3) = \left[P\, \mu_h'(R_1) - 2hQ \cos \frac{2\pi j^*}{N}\, \mu_{h-1}'(R_1)\right], \tag{65}$$

where $P = \mu_h'(V_1)/\mu_h'(V_3) = (N - 3 + 2h)/(N - 3)$, $Q = \mu_{h-1}'(V_1)/\mu_h'(V_3) = 2/(N - 3)$, and $h = 1, 2, \cdots, N$.

6.3. *Approximate moment generating function of C and V when $\rho \neq 0$.* To obtain an approximate moment generating function for C and V when $\rho \neq 0$, we utilize an approximation method given by Leipnik [10]. The exact moment generating function (58) with $\sigma^2 = 1$ can be written as

$$\phi(t_0, t) = (1 - \rho^N)\theta \exp.\left\{-\tfrac{1}{2} \sum_{i=1}^{N} \log\left[1 + \rho^2 - 2t_0 - 2(\rho + t) \cos \frac{2\pi i}{N}\right]\right\}, \tag{66}$$

where $\theta = \prod_{j'}[1 + \rho^2 - 2t_0 - 2(\rho + t)\lambda_{1,j'}]^{\frac{1}{2}}$, and j' refers to the K' fitted terms in m_i. If the sum in the exponent of (66) is replaced by

$$\int_0^N \log\left[1 + \rho^2 - 2t_0 - 2(\rho + t) \cos \frac{2\pi x}{N}\right] dx, \tag{67}$$

and if $(1 - \rho^N)$ is replaced by 1, we obtain the approximate moment generating function

$$\tilde{\phi} = \frac{\prod_{j'} [1 + \rho^2 - 2t_0 - 2(\rho + t)\, \lambda_{1,j'}]^{\frac{1}{2}}}{[\frac{1}{2}(1 + \rho^2 - 2t_0 + \sqrt{(1 + \rho^2 - 2t_0)^2 - 4(\rho + t)^2})]^{\frac{1}{2}N}}. \tag{68}$$

7. Approximate distributions of R.

7.1. *The Pearson Type* I (*Incomplete Beta*) *distribution.* The significance points of ${}_LR$ can be found exactly from equation (48) for $L = 1$ and by integrating equations (52), (53), and (54) for $N = 2L$, $3L$, and $4L$, respectively. These exact probability integrals for $N = 2L$, $3L$, and $4L$ are simply sums of Incomplete Beta functions, and the significance points can be found in Pearson's *Tables of*

the Incomplete Beta-Function [14] or in the Thompson tables [16]. However, the computation of the exact significance points for $L = 1$ and $N > 4$ by use of equation (48) is quite tedious and actually impossible for large N with present logarithm tables and readily available computing devices. Hence, approximate distributions are called for.

The Type I approximation to the distribution of R is

$$f_1(R) = \frac{(1 + R)^{p-1} (1 - R)^{q-1}}{2^{p+q-1} \beta(p, q)}, \; -1 \leq R \leq 1, \tag{69}$$

where p and q are chosen so that the first two moments of this approximate distribution agree with the first two moments of the exact distribution. It can be shown that each moment of the approximate distribution approaches the corresponding exact moment quite rapidly as N increases. On the basis of the approximation, the probability α of the significance point R' being exceeded can be found from the Incomplete Beta function. Thus

$$\alpha = Pr\{R > R'\} = 1 - I_x(p, q) = I_{x'}(p', q'), \tag{70}$$

where

$$I_x(p, q) = \frac{1}{\beta(p, q)} \int_0^x y^{p-1} (1 - y)^{q-1} \, dy, \tag{71}$$

and $x = (1 + R')/2$, $x' = (1 - x)$, $p' = q$, and $q' = p$. Hence, $R' = 2x - 1 = 1 - 2x'$.

The parameters in (69) are taken to be

$$2p = (1 + \mu_1')(1 - \mu_2')/\mu_2, \qquad 2q = (1 - \mu_1')(1 - \mu_2')/\mu_2, \tag{72}$$

where $\mu_2 = \mu_2' - (\mu_1)^2$ and $\mu_i' = \mu_i'(R)$ given in (60). Hence, when the distribution of R is symmetric, $\mu_1' = 0$ and $2p = 2q = (1 - \mu_2')/\mu_2'$.

In Section 3.1, we set up significance points for four special trends for which $\mu_1' = 0$:

(b) $P = 2$; (c) $P = 2, 4$; (d) $P = 2, 3, 6$; (e) $P = 2, 12/5, 3, 4, 6, 12$.
The values of μ_2' for these four trends are:
(b) $(N - 4)/[N(N - 2)]$, (c) $1/(N - 2)$, (d) $1/(N - 4)$, (e) $1/(N - 10)$.
Naturally the third moments for these symmetric distributions are 0. The fourth moments are as follows:

Trend	(b)	(c)	(d)	(e)
Exact	$\frac{3(N^2 - 2N - 10)}{(N + 4)(N + 2)N(N - 2)}$	$\frac{3(N^2 - 2N - 16)}{(N + 2)(N)(N - 2)(N - 4)}$	$\frac{3}{(N - 2)(N - 4)}$	$\frac{3}{(N - 8)(N - 10)}$
Incomplete Beta	$\frac{3(N - 4)^2}{N(N - 2)(N^2 - 8)}$	$\frac{3}{N(N - 2)}$	$\frac{3}{(N - 2)(N - 4)}$	$\frac{3}{(N - 8)(N - 10)}$

We note that for (d) and (e), the fourth moments for the Incomplete Beta are exact and for (b) and (c), they approach the exact values quite rapidly as N increases.

In Section 3.2, we considered some significance points for the following single-period trends: $P = 3, 4, 6$, and 12 The values of $2p$ and $2q$ for these asymmetrical cases are

$$(73) \qquad 2p = \frac{(N - 4 - 2\lambda)E}{D}; \qquad 2q = \frac{(N - 2 + 2\lambda)E}{D},$$

where $\lambda = \cos \frac{2\pi}{P}$, $E = (N - 1)(N - 4) - 4\lambda$ and $D = (N - 3)(N - 1 + 4\lambda) - (N - 1)(1 + 2\lambda)^2$.

Equation (69) has the drawback of using the range $(-1, +1)$ instead of the true range of R, which varies between the last (smallest) λ_k'' to the first (largest) λ_k''. For example, if $N = 12$ and we fit the constant, $\cos \frac{2\pi i}{12}$, and $\sin \frac{2\pi i}{12}$, then $\lambda_{1,1} = 1$, $\lambda_{1,2} = \lambda_{1,3} = \cos \frac{2\pi}{12} = \frac{\sqrt{3}}{2}$, and the range of R is $\left(-1, \cos \frac{4\pi}{12} = \frac{1}{2}\right)$. However, if we fit the constant and $\cos \pi i = (-1)^i$, then $\lambda_{1,1} = 1$ and $\lambda_{1,2} = -1$, the true range would be $\left(-\frac{\sqrt{3}}{2}, +\frac{\sqrt{3}}{2}\right)$. From these examples we see that the error in using the approximate range $(-1, +1)$ varies according to the fitted terms in m_i, and that the error is worse on one tail than on the other, unless symmetric terms are fitted A more accurate approximation could be obtained by use of the exact curtailed range, but it was not thought desirable because the exact range rapidly approaches the approximate range as N increases.

We might add that the significance point, R', can also be calculated from the Inverted Beta (F) distribution, for which tables are given by Merrington and Thompson [13], Snedecor [15], and Fisher and Yates [6]. Cochran [4] has provided an approximate formula for $z = \frac{1}{2} \log_e F$ when n_1 and n_2 are not given in the F-tables.

7.2 *The normal approximation* It should be noted that R is asymptotically normally distributed for $\rho = 0$, as shown by the form of the characteristic function. We have considered the normal approximation with mean $\mu_1'(R)$ and variance $\mu_2(R)$ The variance of R was given in the previous section for the four special trends. For all single period trends, except $P = 2$, $\mu_1' = -(1 + 2\lambda)/(N - 3)$ and the variance is

$$(74) \qquad \mu_2 = \frac{(N - 1 + 4\lambda)}{(N-1)(N-3)} - (\mu_1')^2,$$

where, as before, $\lambda = \cos (2\pi/P)$. Further terms in an asymptotic expansion of the distribution would take account of higher moments of R as Hsu has done for the case of fitting only the mean (m_i = a constant) [7].

REFERENCES

[1] R. L ANDERSON, *Serial correlation in the analysis of time series*, unpublished thesis, Library, Iowa State College, 1941.

[2] R L ANDERSON, "Distribution of the serial correlation coefficient," *Annals of Math. Stat*, Vol 13 (1942), pp. 1–13.

[3] T. W. ANDERSON, "On the theory of testing serial correlation," *Skandinavisk Aktuarietidskrift*, Vol 31 (1948), pp. 88–116.

[4] W. G. COCHRAN, "Note on an approximate formula for the significance levels of z," *Annals of Math. Stat.*, Vol. 11 (1940), pp 93–95

[5] W. J. DIXON, "Further contributions to the problem of serial correlation," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 119–144.

[6] R A. FISHER AND F. YATES, *Statistical Tables for Biological, Agricultural, and Medical Research*, 2d ed., Oliver and Boyd Ltd., 1943.

[7] P. L. HSU, "On the asymptotic distributions of certain statistics used in testing the independence between successive observations from a normal population," *Annals of Math. Stat.*, Vol 17 (1946), pp 350–354.

[8] T. KOOPMANS, "Serial correlation and quadratic forms in normal variables," *Annals of Math. Stat.*, Vol 13 (1942), pp. 14–23.

[9] E. L. LEHMANN, "On optimum tests of composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 18 (1947), p 481

[10] R. B. LEIPNIK, "Distribution of the serial correlation coefficient in a circularly correlated universe," *Annals of Math. Stat*, Vol. 18 (1947), pp. 80–87.

[11] W. G. MADOW, "Contributions to the theory of multivariate statistical analysis," *Trans. Am. Math Soc*, Vol. 44 (1938), p. 461.

[12] W. G. MADOW, "Note on the distribution of the serial correlation coefficient," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 308–310

[13] M. MERRINGTON AND C M. THOMPSON, "Tables of percentage points of the inverted Beta (F) distribution," *Biometrika*, Vol 13 (1943), pp. 73–88.

[14] K. PEARSON, *Tables of the Incomplete Beta-Function*, Cambridge Univ. Press, 1934.

[15] G. W. SNEDECOR, *Statistical Methods*, 4th ed., Iowa State College Press, 1946, pp. 222–225.

[16] C. M. THOMPSON, "Tables of percentage points of the incomplete beta-function," *Biometrika*, Vol. 32 (1941), pp. 151–181.

[17] UNITED STATES DEPARTMENT OF AGRICULTURE, *Agricultural Statistics*, United States Government Printing Office, Washington, D. C , 1939, p. 390.

BAYES SOLUTIONS OF SEQUENTIAL DECISION PROBLEMS

BY A. WALD AND J. WOLFOWITZ

Columbia University

Summary. The study of sequential decision functions was initiated by one of the authors in [1]. Making use of the ideas of this theory the authors succeeded in [4] in proving the optimum character of the sequential probability ratio test. In the present paper the authors continue the study of sequential decision functions, as follows:

a) The proof of the optimum character of the sequential probability ratio test was based on a certain property of Bayes solutions for sequential decisions between two alternatives, the cost function being linear. This fundamental property, the convexity of certain important sets of a priori distributions, is proved in Theorem 3.9 in considerable generality. The number of possible decisions may be infinite.

b) Theorem 3.10 and section 4 discuss tangents and boundary points of these sets of a priori distributions.

(These results for finitely many alternatives were announced by one of us in an invited address at the Berkeley meeting of the Institute of Mathematical Statistics in June, 1948)[1]

c) Theorem 3.6 is an existence theorem for Bayes solutions. Theorem 3.7 gives a necessary and sufficient condition for a Bayes solution. These theorems generalize and follow the ideas of Lemma 1 of [4]

d) Theorems 3.8 and 3.8.1 are continuity theorems for the average risk function. They generalize Lemma 3 in [4]

e) Other theorems give recursion formulas and inequalities which govern Bayes solutions.

1. Introduction. In a previous publication of one of the authors [1] the decision problem was formulated as follows: Let $X = \{x_i\}$ ($i = 1, 2, \cdots$, ad inf.) be a sequence of chance variables. An observation on X is given by a sequence $x = \{x_i\}$ ($i = 1, 2, \cdots$, ad inf.) of real values, where x_i denotes the observed value of X_i. A sequence x is also called a sample or sample point, and the totality M of all possible sample points x is called the sample space. Let $G(x)$ denote the probability that $X_i < x_i$ for $i = 1, 2, \cdots$, ad inf.; i.e., G is the cumulative distribution function of X. In a statistical decision problem G is assumed to be unknown. It is merely known that G is an element of a given class Ω of distribution functions. There is given, furthermore, a space D^* whose elements d represent the possible decisions that can be made in the problem under consideration.

[1] A brief statement of some of the results of the present paper is to be found in the authors' paper of the same name in the *Proc. Nat. Acad. Sci. U. S. A.*, Vol. 35 (1949), pp. 99–102.

The problem is to construct a function $d = D(x)$, called the decision function, which associates with each sample point x an element d of D^* so that the decision $d = D(x)$ is made when x is observed.

Occasionally we shall use the symbol D to denote a decision function $D(x)$. This will be done especially when we want to emphasize that we mean the whole decision function and not merely a particular value of it corresponding to some particular x.

If $d = D(x)$ is the decision function adopted and if $x^0 = \{x_i^0\}$ $(i = 1, 2, \cdots)$ is the particular sample point observed, the number of components of x^0 we have to observe in order to reach a decision is equal to the smallest positive integer $n = n(x^0)$ with the property that $D(x) = D(x^0)$ for any x for which $x_1 = x_1^0, \cdots, x_n = x_n^0$. If no finite n exists with the above property, we put $n(x) = \infty$. If $d(x)$ is equal to a constant d, we put $n(x) = 0$. We shall call $n(x)$ the number of observations required by D when x is the observed sample. Of course, $n(x)$ depends also on the decision rule D adopted. To put this in evidence, we shall occasionally write $n(x, D)$ instead of $n(x)$. If D_0 is a decision function such that $n(x, D_0)$ has a constant value over the whole sample space M, we have the classical non-sequential case. If $n(x, D_0)$ is not constant, we shall say that D_0 is a sequential decision function.

In the remainder of this section we shall sketch briefly some of the fundamental notions of the theory without regard to regularity conditions. The latter will be discussed in the next section.

In [1] a weight function $W(G, d)$ was introduced which expresses the loss suffered by the statistician when G is the true distribution of X and the decision d is made. Let $c(n)$ denote the cost of making n observations; i.e., $c(n)$ is the cost of observing the values of $X_1, \cdots, X_n$. Then, if the decision function $d = D(x)$ is adopted and G is the true distribution of X, the expected value of the loss due to possible erroneous decisions plus the expected cost of experimentation is given by

$$r(G, D) = \int_M W[G, D(x)]\, dG(x) + \int_M c[n(x, D)]\, dG(x). \tag{1.1}$$

The above expression is called the risk when D is the decision function adopted and G is the true distribution.

Let ξ be an a priori probability distribution on Ω; i.e., ξ is a probability measure defined over a suitably chosen Borel field[2] of subsets of Ω. Then the expected value of $r(G, D)$ is given by

$$r(\xi, D) = \int_\Omega r(G, D)\, d\xi. \tag{1.2}$$

[2] A Borel field is an aggregate of sets such that a) the null set is a member of the field, b) the complement with respect to the entire space (here M) is a member of the field, c) the sum of denumerably many members of the field is itself in the field.

The above expression is called the risk when ξ is the a priori distribution on Ω and D is the decision function adopted.

We shall say that the decision function D_0 is a Bayes solution relative to the a priori distribution ξ if

$$r(\xi, D_0) \leqq r(\xi, D) \text{ for all } D. \tag{1.3}$$

If there existed an a priori distribution on Ω and if this distribution were known, we could put ξ equal to this a priori distribution and a Bayes solution relative to ξ would provide a very satisfactory solution of the decision problem. In most applications, however, not even the existence of an a priori distribution can be postulated. Nevertheless, the study of Bayes solutions corresponding to various a priori distributions is of great interest in view of some results given in [1]. It was shown in [1] that under rather general conditions the class C of the Bayes solutions corresponding to all possible a priori distributions ξ has the following property: If D_1 is a decision function that is not an element of C, there exists a decision function D_2 in C such that

$$r(G, D_2) \leqq r(G, D_1) \text{ for all } G \tag{1.4}$$

and

$$r(G, D_2) < r(G, D_1) \text{ for at least one } G. \tag{1.5}$$

It was furthermore shown in [1] that under general conditions a minimax solution D_0 of the decision problem is also a Bayes solution corresponding to some a priori distribution ξ. By a minimax solution we mean a decision function D_0 such that, for all D

$$\operatorname*{Sup}_G r(G, D_0) \leqq \operatorname*{Sup}_G r(G, D). \tag{1.6}$$

2. Regularity conditions and other assumptions. We shall make the following assumptions:

Assumption 1. *The chance variables are identically and independently distributed. The common distribution is either discrete or absolutely continuous.*

Let $p(a \mid F)$ denote the elementary probability law of X_i when F is the distribution of X_i; i.e., when F is discrete, $p(a \mid F)$ is the probability that $X_i = a$, and when F is absolutely continuous, $p(a \mid F)$ is the probability density of X_i at a.

In the space M of sequences x let B be the smallest Borel field which contains all sets of points x which are defined by the relations

$$x_i < a_i \qquad i = 1, 2, \cdots \text{ ad inf.},$$

where the a_i are real numbers or $+\infty$. Each admissible[3] F induces a probability measure $F^*(B)$ on M; the totality of these probability measures is Ω. Let H^*

[3] An F or F* is admissible if F* is in Ω.

be a given Borel field of subsets of Ω. The only subsets of Ω which we shall discuss in this paper will be members of H^*, and all probability measures on Ω which we shall discuss will be measurable (H^*). This will henceforth be assumed without further repetition.

Let A^* be any set in H^*, and A the set of F which corresponds to the F^* in A^*. The sets A form a Borel field, say H. By definition, the probability measure of a set A according to a probability measure $\xi(H^*)$ on Ω is to be the same as the probability measure of A^* according to ξ.

Let $M \times \Omega$ be the Cartesian product of M and Ω ([5], page 82), and K be the smallest Borel field of subsets of $M \times \Omega$ which contains the Cartesian product of any member of B by any member of H^*.

For a given decision function $d = D(x)$, $W(F, D(x))$ is a function of F and x. Hereafter in this paper we shall limit ourselves to functions $D(x)$ such that $W(F, D(x))$ is measurable (K), and $n(x, D)$ is measurable (B).

It is true that in Section 1, W was given as a function of G, the distribution of X. Because of Assumption 1, $G = F^*$, and there is a one-to-one correspondence between F and F^*. Thus we may, in appropriate places, interchange them freely.

ASSUMPTION 2. *For every real a, except possibly on a Borel set*[4] *whose probability is zero according to every admissible F, $p(a \mid F)$ exists and is a function of a and F which is measurable (K). If the admissible distributions F are discrete, there exists a fixed sequence $\{b_i\}$ ($i = 1, 2, \cdots$, ad inf.) of real values such that $\sum_{i=1}^{\infty} p(b_i \mid F) = 1$ for all admissible F.*

ASSUMPTION 3. *$W(F, d)$ is bounded. For every d in D^*, $W(F, d)$ is a function of F which is measurable (H).*

In what follows ξ will always denote a probability measure (H^*) on Ω. Thus

$$W(\xi, d) = \int_\Omega W(F, d)\, d\xi$$

exists.

ASSUMPTION 4. *The function $c(n) = cn$. Without loss of generality we may take $c = 1$, so that $c(n) = n$.*

We shall introduce the following convergence definition in the space D^*: the sequence $\{d_i\}$ converges to d_0 if

$$\lim_{i\to\infty} W(F, d_i) = W(F, d_0)$$

uniformly in the admissible F's.

ASSUMPTION 5. *The space D^* is compact in the sense of the above convergence definition.*

One can easily verify that, if $\lim_{i\to\infty} d_i = d_0$, then

$$\lim_{i\to\infty} W(\xi, d_i) = W(\xi, d_0);$$

[4] A Borel set is a member of the smallest Borel field which contains all the open sets of the real line.

i.e., $W(\xi, d)$ is a continuous function of d. Thus, because of Assumption 5, the minimum of $W(\xi, d)$ with respect to d exists.

We shall now show that, under the above conditions

$$\int_M W[F^*, D(x)]\, dF^*(x) \tag{2.1}$$

exists and is a function of F^* measurable (H^*). For any j let R_j be the set in B such that $n(x, D) = j$. Then it is enough to show that, for any j,

$$\int_{R_j} W[F^*, D(x)]\, dF^*(x) \tag{2.2}$$

exists and is a function of F^* measurable (H^*).

In the discrete case, the integral (2.2) is equal to the sum[5]

$$\sum_{(x_1, \dots, x_j) \in R_j} W[F^*, D(x)] p(x_1 \mid F) \cdots p(x_j \mid F). \tag{2.3}$$

For fixed values of $x_1, \cdots, x_j$, the expression under the summation sign is obviously a function of F^* measurable (H^*). Since, because of Assumption 2, there are only countably many points $(x_1, \cdots, x_j)$ in R_j, the sum (2.3) must be a function of F^* measurable (H^*).

In the absolutely continuous case, the integral (2.2) is equal to (2.4)

$$\int_{R_j} W[F^*, D(x)] \prod_{i=1}^{j} p(x_i \mid F)\, d\nu(j) \tag{2.4}$$

where $\nu(j)$ is Borel measure in the j-dimensional Euclidean space. The integrand is measurable (K). Hence, the integral (2.4) exists and is a function of F^* measurable (H^*) (see [5], Chapter III, Theorems 9.3 and 9.8).

3. Some results concerning Bayes solutions. If ξ is the a priori probability measure on Ω, the a posteriori probability of a subset ω of Ω for given values $x_1, \cdots, x_m$ of the first m chance variables is given by

$$\xi(\omega \mid \xi, x_1, \cdots, x_m) = \frac{\int_\omega p(x_1 \mid F) \cdots p(x_m \mid F)\, d\xi}{\int_\Omega p(x_1 \mid F) \cdots p(x_m \mid F)\, d\xi}. \tag{3.1}$$

Let

$$\rho_0(\xi) = \operatorname*{Min}_d W(\xi, d). \tag{3.2}$$

For any positive integral value m, let $\rho_m(\xi)$ denote the infimum of $r(\xi, D)$ with respect to D where D is restricted to decision functions for which $n(x, D) \leqq m$ for all x. For any positive integer m, let $d = D^m(x)$ denote a decision function

[5] Because of the definition of R_j we may, in the expressions (2.3) and (2.4), proceed as if R_j were a Borel set in j-dimensional Euclidean space.

D for which $n(x, D) \leqq m$ for all x. Thus, we can write

$$(3.3) \qquad \rho_m(\xi) = \operatorname*{Inf}_{D^m} r(\xi, D^m) \quad (m = 1, 2, \cdots, \text{ad inf.}).$$

Let

$$(3.4) \qquad \rho(\xi) = \operatorname*{Inf}_{D} r(\xi, D).$$

We shall first prove several theorems concerning the functions $\rho_0(\xi)$, $\rho_m(\xi)$, and $\rho(\xi)$.

THEOREM 3.1. *The following recursion formula holds:*[6]

$$(3.5) \qquad \rho_{m+1}(\xi) = \operatorname{Min}\left[\rho_0(\xi), 1 + \int_{-\infty}^{\infty} \rho_m(\xi_a)\, p(a \mid \xi)\, da\right]$$
$$(m = 0, 1, 2, \cdots, \text{ad inf.})$$

where

$$(3.6) \qquad \xi_a(\omega) = \xi(\omega \mid \xi, a) \text{ and } p(a \mid \xi) = \int_\Omega p(a \mid F)\, d\xi.$$

PROOF: Let $\rho_m^*(\xi)$ $(m = 1, 2, \cdots, \text{ad inf.})$ denote the infimum of $r(\xi, D)$ with respect to D where D is subject to the restriction that $n(x, D) \geqq 1$ and $\leqq m$ for all x. Clearly,

$$(3.7) \qquad \rho_{m+1}(\xi) = \operatorname{Min}[\rho_0(\xi), \rho_{m+1}^*(\xi)].$$

Let $\rho_m^*(\xi \mid a)$ denote the infimum with respect to D of the conditional risk (conditional expected value of $W[F, D(x)] + n(x, D)$) when the first observation x_1 on X_1 is a and D is restricted to decision functions for which $n(x, D) \geqq 1$ and $\leqq m$ for all x. Let $\bar{D}(m)$ be the temporary generic designation of such a decision function. Let $\bar{D}(m \mid a)$ be the decision function which is obtained from $\bar{D}(m)$ when the first observation is a. Finally let $r(\xi, D \mid a)$ be the conditional risk when the a priori distribution function is ξ, D is the decision function and requires at least one observation, and the first observation is a. We then have that

$$r(\xi, \bar{D}(m + 1) \mid a) = r(\xi_a, \bar{D}(m + 1 \mid a)) + 1.$$

Hence

$$(3.8) \qquad \rho_{m+1}^*(\xi \mid a) = \rho_m(\xi_a) + 1.$$

The unconditional quantity $\rho_{m+1}^*(\xi)$ must clearly be equal to the average value of the infimum of the conditional risk. Thus we have

$$(3.9) \qquad \rho_{m+1}^*(\xi) = \int_{-\infty}^{\infty} \rho_{m+1}^*(\xi \mid a) p(a \mid \xi) da.$$

[6] If the distribution of X is discrete, the integration with respect to a is to be replaced by summation with respect to a. This remark refers also to subsequent formulas.

Equation (3.5) follows from (3.7), (3.8) and (3.9).

THEOREM 3.2. *The function $\rho(\xi)$ satisfies the following equation:*

$$\rho(\xi) = \operatorname{Min}\left[\rho_0(\xi), \int_{-\infty}^{\infty} \rho(\xi_a)p(a \mid \xi)\, da + 1\right]. \tag{3.10}$$

The proof of this theorem is omitted, since it is essentially the same as that of Theorem 3.1.

THEOREM 3.3.[7] *The following inequalities hold:*

$$0 \leqq \rho_m(\xi) - \rho(\xi) \leqq \frac{W_0^2}{m} \qquad (m = 1, 2, \cdots, \text{ad inf.}) \tag{3.11}$$

where W_0 is the least upper bound of $W(F, d)$.

PROOF: Let $\{D_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ be a sequence of decision functions such that

$$\lim_{i=\infty} r(\xi, D_i) = \rho(\xi). \tag{3.12}$$

Let, furthermore, $P_i(\xi)$ denote the probability that at least m observations will be made when D_i is the decision function adopted and ξ is the a priori probability measure on Ω. Since $\rho(\xi) \leqq W_0$ and since

$$r(\xi, D_i) \geqq mP_i(\xi), \tag{3.13}$$

it follows from (3.12) that

$$\limsup_{i=\infty} P_i(\xi) \leqq \frac{W_0}{m}. \tag{3.14}$$

Let D_i^m be the decision function obtained from D_i as follows: $D_i^m(x) = D_i(x)$ for all x for which $n(x, D_i) \leqq m$. $D_i^m(x)$ is equal to a fixed element d_0 for all x for which $n(x, D_i) > m$.[8]

Clearly,

$$r(\xi, D_i^m) \leqq r(\xi, D_i) + P_i(\xi)W_0. \tag{3.15}$$

From (3.12), (3.14) and (3.15) it follows that

$$\limsup_{i=\infty} r(\xi, D_i^m) \leqq \rho(\xi) + \frac{W_0^2}{m}. \tag{3.16}$$

Since $\rho_m(\xi)$ cannot exceed the left hand member of (3.16), the second half of (3.11) follows from (3.16). The first half of (3.11) is obvious.

---

[7] This theorem is essentially the same as Lemma 2.1 in [6].

[8] We verify that $W(F, D_i^m)$ is measurable (K), as follows: Consider the set V of couples (F, x) such that $W(F, D_i^m(x)) < c$, where c is some real constant. We want to show that $V \epsilon K$ For this purpose let V_0 be the set of couples (F, x) such that $W(F, D_i(x)) < c$. Then $V_0 \epsilon K$ Let V_1 be the set of x's such that $n(x, D_i) \leqq m$. Then $V_1 \epsilon B$, $(\Omega \times V_1) = V_2 \epsilon K$, $V_0V_2 \epsilon K$ Let $V_3 = M - V_1$ For every $x \epsilon V_3$ we have $W(F, D_i^m(x)) = W(F, d_0)$. Let V_4 be the set of F's such that $W(F, d_0) < c$. Then $V_4 \epsilon H$ by Assumption 3. Finally we have $V = V_0V_2 + V_4 \times V_3$, so that $V \epsilon K$.

The immediate consequence of Theorem 3 3 is the relation[9]

$$\lim_{m\to\infty} \rho_m(\xi) = \rho(\xi). \tag{3.17}$$

THEOREM 3.4. *If ξ_1 and ξ_2 are two probability measures on Ω such that*[10]

$$\frac{\xi_1(\omega)}{\xi_2(\omega)} \leqq 1 + \epsilon \text{ for all } \omega, \tag{3.18}$$

then

$$\rho(\xi_1) \leqq (1 + \epsilon)\rho(\xi_2). \tag{3.19}$$

PROOF: It follows from (3.18) that

$$r(\xi_1, D) \leqq (1 + \epsilon)r(\xi_2, D) \text{ for all } D. \tag{3.20}$$

Hence, (3.19) must hold.

The above theorem permits the computation of a simple and in many cases useful lower bound of $\int_{-\infty}^{\infty} \rho(\xi_a)p(a \mid \xi)\, da$ as follows:

For any real value a, let ϵ_a be a non-negative value (not necessarily finite) determined such that

$$\frac{\xi(\omega)}{\xi_a(\omega)} \leqq 1 + \epsilon_a \text{ for all } \omega. \tag{3.21}$$

Then

$$\int_{-\infty}^{\infty} \rho(\xi_a)\, p(a \mid \xi)\, da \geqq \int_{-\infty}^{\infty} \frac{\rho(\xi)}{1 + \epsilon_a}\, p(a \mid \xi)\, da = \rho(\xi) \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da. \tag{3.22}$$

Since $\epsilon_a \geqq 0$ and since $\rho_0(\xi) \geqq \rho(\xi)$, we obviously have

$$\rho(\xi) \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da \geqq \rho(\xi) - \left[1 - \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da\right]\rho_0(\xi). \tag{3.23}$$

Hence, we obtain the inequality

$$\int_{-\infty}^{\infty} \rho(\xi_a)\, p(a \mid \xi)\, da \geqq \rho(\xi) - \rho_0(\xi)\left[1 - \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da\right]. \tag{3.24}$$

An upper bound of the left hand member in (3.24) is obtained by replacing ρ by ρ_0; i.e.,

$$\int_{-\infty}^{\infty} \rho(\xi_a)p(a \mid \xi)\, da \leqq \int_{-\infty}^{\infty} \rho_0(\xi_a)p(a \mid \xi)\, da. \tag{3.25}$$

---

[9] A proof of (3.17) is contained implicitly in the work of Arrow, Blackwell and Girshick ([2], Section 1.3).

[10] The left member of (3.18) is defined to be equal to 1 when $\xi_1(\omega) = \xi_2(\omega) = 0$.

The bounds given in (3.24) and (3.25) may be useful in constructing Bayes solutions, since the following theorem holds:

THEOREM 3.5. *If*

$$\rho_0(\xi) > \int_{-\infty}^{\infty} \rho_0(\xi_a) p(a \mid \xi)\, da + 1, \tag{3.26}$$

then $\rho(\xi) < \rho_0(\xi)$. *If*

$$\rho_0(\xi)\left[1 - \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da\right] < 1, \tag{3.27}$$

then $\rho(\xi) = \rho_0(\xi)$.

The above theorem is an immediate consequence of (3.10), (3.24) and (3.25).

A decision procedure relative to a given a priori probability measure ξ_0 will be given with the help of the function $\rho(\xi)$ as follows: If $\rho(\xi_0) = \rho_0(\xi_0)$, take a final decision d for which $W(\xi_0, d)$ is minimized. If $\rho(\xi_0) < \rho_0(\xi_0)$, take an observation on X_1 and compute the a posteriori probability measure ξ_1. If $\rho(\xi_1) = \rho_0(\xi_1)$, stop experimentation with a final decision d for which $W(\xi_1, d)$ is minimized If $\rho(\xi_1) < \rho_0(\xi_1)$, take an observation on X_2 and compute the a posteriori probability measure ξ_2 corresponding to the observed values of X_1 and X_2, and so on. The above decision procedure will be shown later to be a Bayes solution. Theorem 3 5 permits one to decide whether $\rho(\xi) < \rho_0(\xi)$ or $= \rho_0(\xi)$ whenever ξ satisfies (3.26) or (3.27). Theorem 3.5 will be useful when the class of all ξ's for which neither (3.26) nor (3.27) holds is small.

For the purposes of the next theorem let $\hat{D}$ designate the decision procedure described in the preceding paragraph. (We shall shortly show that $\hat{D}$ is a decision function in the sense of our definition.)

Let $\hat{D}^0$ be the decision procedure where the first observation is taken and then one proceeds according to $\hat{D}$.

We shall now prove that $\hat{D}$ and $\hat{D}^0$ are Bayes solutions. More precisely, we shall prove the following theorem:[11]

THEOREM 3.6. *For any ξ, $\hat{D}$ and $\hat{D}^0$ as defined above are decision functions. Let D be any decision function for which $n(x, D) \geqq 1$ and let*

$$\rho^*(\xi) = \operatorname*{Inf}_{D} r(\xi, D).$$

Then

$$r(\xi, \hat{D}) = \rho(\xi)$$

and

$$r(\xi, \hat{D}^0) = \rho^*(\xi).$$

[11] This theorem follows also from some earlier more general existence theorems ([6], Theorems 2.4 and 3.3). (See also [4], Lemma 1) The validity of Theorem 3 6 was proved also by Arrow, Blackwell and Girshick [2].

In view of this theorem, the operation "infimum with respect to D" in the definitions of $\rho(\xi)$, and $\rho^*(\xi)$ can be replaced by "minimum with respect to D."

First we shall establish the measurability properties of $\hat{D}$ and $\hat{D}^0$. Since the proofs are similar, we restrict ourselves to consideration of $\hat{D}$. Let $\xi_{x_1,\dots,x_m}$ be the a posteriori distribution (3.1). From the (B) measurability of $\rho_0(\xi_{x_1,\dots,x_m})$ and $\rho(\xi_{x_1,\dots,x_m})$ it follows easily that $n(x, \hat{D})$ is measurable (B). It remains to prove that $W(F, \hat{D}(x))$ is measurable (K). For this purpose, let $L^i = (d_1^i, \cdots, d_{k_i}^i)$ be a sequence $\frac{1}{i}$ dense in D^*, i.e., for any $d \in D^*$ there exists a $g \in D^*$ such that $g \in L^i$ and $| W(F, d) - W(F, g) | < \frac{1}{i}$ uniformly in F. (The existence of such a sequence follows from Assumption 5.) Let now $D_i(x)$ be a decision function defined as follows:

$$n(x, D_i) = n(x, \hat{D}).$$

Suppose $n(x, \hat{D}) = m$ when the observations are $x_1, \cdots, x_m$. We define $D_i(x)$ to be such that $D_i(x)$ is an element of L^i and

$$W(\xi_{x_1,\dots,x_m}, D_i(x)) = \underset{d \in L^i}{\mathrm{Min}}\, W(\xi_{x_1,\dots,x_m}, d), \tag{3.28}$$

i.e., $D_i(x)$ takes the minimizing value of d. For any fixed d, the set of x's satisfying the equation $D_i(x) = d$ is without difficulty shown to be (B) measurable. Since $D_i(x)$ assumes only a finite number of values in D^*, it follows from Assumption 3 that $W(F, D_i(x))$ is measurable (K). Now

$$\lim_{i\to\infty} W(F, D_i(x)) = W(F, \hat{D}(x)),$$

so that $W(F, \hat{D}(x))$ is measurable (K).

We shall now prove that $\hat{D}$ is a Bayes solution, i.e., that

$$\rho(\xi) = r(\xi, \hat{D}). \tag{3.29}$$

In a similar way it can be proved that

$$\rho^*(\xi) = r(\xi, \hat{D}^0). \tag{3.30}$$

If $\rho_0(\xi) = \rho(\xi)$, there can be no better decision function (from the point of view of reducing the risk) than $\hat{D}$, i.e., $\hat{D}$ is a Bayes solution. Suppose then that

$$\rho_0(\xi) > \rho(\xi). \tag{3.31}$$

If (3.31) holds and $\hat{D}$ is not a Bayes solution, there exists a decision function $\bar{D}_1$ such that

$$r(\xi, \bar{D}_1) < r(\xi, \hat{D}) \tag{3.32}$$

and

$$r(\xi, \bar{D}_1) < \frac{\rho_0(\xi) + \rho(\xi)}{2}. \tag{3.33}$$

Now $\bar{D}_1$ must require that at least one observation be taken, else (3.33) could not hold. Thus $\hat{D}$ and $\bar{D}_1$ both require at least one observation

Suppose one observation is taken. Let $r(\xi, D \mid a)$ denote the conditional risk of proceeding according to D when ξ is the a priori distribution and a is the first observation. For a given D we have that $r(\xi, D \mid a)$ is a function only of ξ_a. In particular $r(\xi, \hat{D} \mid a)$ and $r(\xi, \bar{D}_1 \mid a)$ are functions only of ξ_a.

We can now apply to $r(\xi, \hat{D} \mid a)$ and $r(\xi, \bar{D}_1 \mid a)$ the same argument that was applied above to $r(\xi, \hat{D})$ and $r(\xi, \bar{D}_1)$, and conclude again as follows: whenever $\rho_0(\xi_a) = \rho(\xi_a)$ (when one takes no more observations according to $\hat{D}$), taking additional observations cannot diminish the conditional risk below $r(\xi, \hat{D} \mid a)$ ($\bar{D}_1$ may require an additional observation without having

$$r(\xi, \bar{D}_1 \mid a) > r(\xi, \hat{D} \mid a).$$

This can happen when $\rho_0(\xi_a) = \rho^*(\xi_a)$). Whenever $\rho_0(\xi_a) > \rho(\xi_a)$ (when $\hat{D}$ requires us to take another observation) two cases may occur: either a) $\bar{D}_1$ requires us to take another observation, in which case its decision is the same as that of $\hat{D}$, or b) $\bar{D}_1$ requires us to stop taking observations. There exists then another decision function whose conditional risk is less than

$$\frac{\rho_0(\xi_a) + \rho(\xi_a)}{2} + 1.$$

Both this decision function and $\hat{D}$ require that another observation be taken. We conclude that up to and including the first observation, $\hat{D}$ coincides either with $\bar{D}_1$ or with another decision function $\bar{D}_2$ whose risk is not greater than that of $\bar{D}_1$.

We continue in this manner for 2, 3, $\cdots$ observations. The above argument is always valid because of Assumption 4 and because the past history of the process (the sequence of observations) enters only through the a posteriori probability. Thus we conclude that for any positive integer k there exists a decision function $\bar{D}_k$ such that up to and including the k-th observation $\hat{D}$ gives the same decision as $\bar{D}_k$ and the risk corresponding to $\bar{D}_k$ does not exceed the risk corresponding to $\bar{D}_1$. Since $\lim_{k\to\infty} r(\xi, \bar{D}_k) \geqq r(\xi, \hat{D})$, (3.32) cannot hold. Hence (3.29) holds and $\hat{D}$ is a Bayes solution.

For any probability measure ξ on Ω one of the following three conditions must hold:

(1) $\mathrm{Min}_d\, W(\xi, d) < r(\xi, D)$ for any D for which $n(x, D) \geqq 1$.

(2) $\mathrm{Min}_d\, W(\xi, d) \leqq r(\xi, D)$ for all D for which $n(x, D) \geqq 1$, and the equality sign holds for at least one D with $n(x, D) \geqq 1$.

(3) There exists a D with $n(x, D) \geqq 1$ such that $\mathrm{Min}_d\, W(\xi, d) > r(\xi, D)$.

In view of Theorem 3.6, the conditions (1), (2) and (3) can be expressed by: (1) $\rho_0(\xi) < \rho^*(\xi)$, (2) $\rho_0(\xi) = \rho^*(\xi)$ and (3) $\rho_0(\xi) > \rho^*(\xi)$, respectively.

We shall say that a probability measure ξ on Ω is of the first type if it satisfies (1), of the second type if it satisfies (2), and of the third type if it satisfies (3). Since the a posteriori probability defined in (3.1) is also a probability measure

on Ω, any a posteriori probability measure will be one of the three types mentioned above.

We shall now prove the following characterization theorem:

THEOREM 3.7.[12] *A necessary and sufficient condition for a decision function $d = D_0(x)$ to be a Bayes solution relative to a given a priori distribution ξ_0 is that the following three relations be fulfilled for any sample point x, except perhaps on a set whose probability measure is zero when ξ_0 is the a priori distribution in Ω:*

(a) *For any $m < n(x, D_0)$, the a posteriori distribution $\xi(\omega \mid \xi_0, x_1, \cdots, x_m)$* is *either of the second or of the third type,*

(b) *For $m = n(x, D_0)$, the a posteriori distribution $\xi(\omega \mid \xi_0, x_1, \cdots, x_m)$ is either of the first or the second type,*

(c) *For $m = n(x, D_0)$, we have*

$$\operatorname*{Min}_d W(\xi_{x_1,\ldots,x_m}, d) = W(\xi_{x_1,\ldots,x_m}, D_0(x))$$

where $\xi_{x_1,\ldots,x_m}$ stands for an a priori distribution that is equal to the a posteriori distribution corresponding to $\xi_0, x_1, \cdots, x_m$.

PROOF: We shall omit the proof of the sufficiency of the conditions (a), (b) and (c), since it is essentially the same as that of Theorem 3.6. To prove the necessity of these conditions, let $d = D_0(x)$ be a decision function and let M^* denote the set of all sample points x for which at least one of the relations (a), (b) and (c) is violated. First, we shall show tht M^* is a set measurable (B). Let M_1^* be the set of all x's for which (a) is violated, M_2^* the set of all x's for which (b) is violated, and M_3^* the set of all x's for which (c) is violated. Clearly, M^* is shown to be measurable (B) if we can show that $M_i^* (i = 1, 2, 3)$ is measurable (B). Let $M_{ir}^* (r = 1, 2, \cdots$, ad inf) denote the subset of M_i^* for which the first violation of the corresponding condition occurs for the sample $x_1, \cdots, x_r$. We merely have to show that M_{ir}^* is measurable (B) for all i and r. The measurability of M_{3r}^* follows from the fact that $\operatorname{Min}_d W(\xi_{x_1,\ldots,x_r}, d)$ and

$$W[\xi_{x_1,\ldots,x_r}, D_0(x)]$$

are functions of x measurable (B). To show the measurability of M_{1r}^* and M_{2r}^*, it is sufficient to show that the set of all samples $x_1, \cdots, x_r$ for which $\xi_{x_1,\ldots,x_r}$ is of type $i (i = 1, 2, 3)$ is measurable (B). But this follows from the fact that $\rho_0(\xi_{x_1,\ldots,x_r})$ and $\rho^*(\xi_{x_1,\ldots,x_r})$ are functions of $(x_1, \cdots, x_r)$ measurable (B). Hence, M^* is proved to be measurable (B).

For any x in M^* let $m(x)$ be the smallest positive integer such that at least one of the relations (a), (b) and (c) is violated for the finite sample

$$x_1, x_2, \cdots, x_{m(x)}.$$

Clearly, if x is a point in M^*, then also any sample point y is in M^* for which $y_1 = x_1, \cdots, y_{m(x)} = x_{m(x)}$. Let x^0 be any particular sample point in M^* and let $r(\xi_0, D_0, x_1^0, \cdots, x_{m(x^0)}^0)$ denote the conditional risk when ξ_0 is the a priori

[12] See also the proof of Lemma 1 in [4].

distribution in Ω, D_0 is the decision function adopted and the first $m(x^0)$ observations are equal to $x_1^0, \cdots, x_{m(x^0)}^0$, respectively; i.e., $r(\xi_0, D_0, x_1^0, \cdots, x_{m(x^0)}^0)$ is the conditional expected value of $W(F, D_0(x)) + n(x, D_0)$, when ξ_0 is the a priori distribution in Ω, D_0 is the decision function adopted and $x_1^0, \cdots, x_{m(x^0)}^0$ are the first $m(x^0)$ observations.

Let $D_1(x)$ be the decision function determined as follows: for any x not in M^* we put $D_1(x) = D_0(x)$. For any x in M^*, let $n(x_1, D_1)$ be equal to the smallest integer $n(x) \geqq m(x)$ for which

$$\rho_0(\xi_{x_1, \ldots, x_{n(x)}}) = \rho(\xi_{x_1, \ldots, x_{n(x)}})$$

and the value of $D_1(x)$ is determined so that condition (c) of our theorem is fulfilled. Since, for any positive integer m, the subset of M^* where $m(x) = m$ is (B) measurable, $D_1(x)$ has the proper measurability properties. Applying Theorem 3.6, we see that

$$r(\xi_0, D_1, x_1, \cdots, x_{m(x)}) = \rho(\xi_{x_1, \ldots, x_{m(x)}}) \tag{3.34}$$

for any x in M^*. On the other hand, since D_0 violates at least one of the conditions (a), (b), and (c) at every point x in M^*, we have

$$r(\xi_0, D_0, x_1, \cdots, x_{m(x)}) > \rho(\xi_{x_1, \ldots, x_{m(x)}}) \tag{3.35}$$

for every x in M^*. If the probability measure of M^* is positive when ξ_0 is the a priori probability measure, the above two relations imply that

$$r(\xi_0, D_0) > r(\xi_0, D_1).$$

Thus, D_0 is not a Bayes solution and the proof of Theorem 3.7 is complete.

We shall now prove the following continuity theorem.[13]

THEOREM 3.8. *Let* $\{\xi_i\}$ $(i = 0, 1, 2, \cdots, ad\ inf.)$ *be a sequence of probability measures on* Ω *such that*

$$\lim_{i \to \infty} \frac{\xi_i(\omega)}{\xi_0(\omega)} = 1 \text{ uniformly in } \omega. \tag{3.36}$$

Then

$$\lim_{i \to \infty} \rho(\xi_i) = \rho(\xi_0). \tag{3.37}$$

PROOF: It follows from (3.36) that for any $\epsilon > 0$, we have for almost all values i

$$\frac{\xi_i(\omega)}{\xi_0(\omega)} < 1 + \epsilon \text{ and } \frac{\xi_0(\omega)}{\xi_i(\omega)} < 1 + \epsilon \text{ for all } \omega. \tag{3.38}$$

Our theorem is an immediate consequence of (3.38) and Theorem 3.4.

[13] A proof of this theorem for finite Ω was given by G. W. Brown and is included in [2]. See also Lemma 3 in [4].

A stronger continuity theorem is the following:

THEOREM 3.8.1. *Let* $\{\xi_i\}$, $(i = 0, 1, 2, \cdots, ad\ inf)$ *be a sequence of probability measures on* Ω *such that*

$$\lim_{i\to\infty} \xi_i(\omega) = \xi_0(\omega)$$

uniformly in ω. *Then* (3.37) *holds.*

PROOF: It follows from (3.11) that

$$\lim_{m\to\infty} \rho_m(\xi) = \rho(\xi)$$

uniformly in ξ. Hence it is sufficient to prove that, under the conditions of the theorem,

$$\lim_{i\to\infty} \rho_m(\xi_i) = \rho_m(\xi_0)$$

for any m. Let $D^m(x)$ denote a decision function for which $n(x, D^m) \leqq m$ for all x. It follows that, for a fixed m, $r(F, D^m)$ is bounded, uniformly in F and D^m (Assumptions 3 and 4). From the hypothesis on $\{\xi_i\}$ it then follows that

$$\lim_{i\to\infty} r(\xi_i, D^m) = r(\xi_0, D^m)$$

uniformly in D^m. From this the desired result follows readily.

A class C of probability measures ξ on Ω will be said to be convex if for any two elements ξ_1 and ξ_2 of C and for any positive value $\lambda < 1$, the probability measure $\xi = \lambda\xi_1 + (1 - \lambda)\xi_2$ is an element of C.

For any element d_0 of D, let C_{i,d_0} denote the class of all probability measures ξ of type i $(i = 1, 2, 3)$ for which $W(\xi, d_0) = \operatorname{Min}_d W(\xi, d)$. Let C_d denote the set-theoretical sum of $C_{1,d}$ and $C_{2,d}$. We shall now prove the following theorem.

THEOREM 3.9. *For any element* d, *the classes* $C_{1,d}$ *and* C_d *are convex.*[14]

Let ξ_1 and ξ_2 be two elements of $C_{1,d}$. Then for any decision function $D(x)$ which requires at least one observation we have

$$W(\xi_1, d) < r(\xi_1, D) \text{ and } W(\xi_2, d) < r(\xi_2, d). \tag{3.39}$$

Let $\xi = \lambda\xi_1 + (1 - \lambda)\xi_2$ where λ is a positive number <1. Clearly,

$$W(\xi, d) = \lambda W(\xi_1, d) + (1 - \lambda) W(\xi_2, d) \tag{3.40}$$

and

$$r(\xi, D) = \lambda r(\xi_1, D) + (1 - \lambda) r(\xi_2, D). \tag{3.41}$$

From (3.39), (3.40) and (3.41) we obtain

$$W(\xi, d) < r(\xi, D) \quad \text{and} \quad W(\xi, d) = \operatorname{Min}_{d^*} W(\xi, d^*). \tag{3.42}$$

Hence ξ is an element of $C_{1,d}$ and the convexity of $C_{1,d}$ is proved. The convexity of C_d can be proved in the same way by replacing $<$ by $\leqq$ in (3.39) and (3.42).

[14] See also Lemma 2 in [4].

We shall say that a set L of probability measures ξ is a linear manifold if for any two elements ξ_1 and ξ_2 of L, $\xi = \alpha\xi_1 + (1 - \alpha)\xi_2$ is also an element of L for any real value α for which $\alpha\xi_1 + (1 - \alpha)\xi_2$ is a probability measure. A linear manifold L will be said to be tangent to C_d if the intersection of L and $C_{2,d}$ is not empty, but the intersection of L and $C_{1,d}$ is empty

For any decision function $D(x)$ and for any element d of D^*, let $L(D, d)$ denote the linear manifold consisting of all ξ which satisfy the equation

$$W(\xi, d) = r(\xi, D). \tag{3.43}$$

THEOREM 3.10. *Let ξ_0 be an element of $C_{2,d}$ and let $D_0(x)$ be a decision function that requires at least one observation and is such that $W(\xi_0, d) = r(\xi_0, D_0)$. Then the linear manifold $L(D_0, d)$ is tangent to C_d.*

PROOF: ξ_0 is obviously an element of $L(D_0, d)$. Thus the intersection of $L(D_0, d)$ and $C_{2,d}$ is not empty. For any element ξ_1 of $C_{1,d}$ we have $W(\xi_1, d) < r(\xi_1, D)$ for any D that requires at least one observation. Hence, $W(\xi_1, d) < r(\xi_1, D_0)$ and, therefore, ξ_1 cannot be an element of $L(D_0, d)$. This proves our theorem.

4. Applications to the case where Ω and D^* are finite. In this section we shall apply the general results of the preceding section to the following special case: the space Ω consists of a finite number of elements, $F_1, \cdots, F_k$ (say), and the space D^* consists of the elements $d_1, \cdots, d_k$ where d_i denotes the decision to accept the hypothesis H_i that F_i is the true distribution. Let

$$W(F_i, d_j) = W_{ij} = 0 \text{ for } i = j \text{ and } >0 \text{ for } i \neq j. \tag{4.1}$$

It will be sufficient to discuss the cases $k = 2$ and $k = 3$, since the extension to $k > 3$ will be obvious. We shall first consider the case $k = 2$. In this case any a priori distribution ξ is represented by two numbers g_1 and g_2 where g_i is the a priori probability that F_i is true ($i = 1, 2$). Thus, $g_i \geqq 0$ and $g_1 + g_2 = 1$. Let ξ_i denote the a priori distribution corresponding to $g_i = 1$ ($i = 1, 2$). Clearly C_{d_1} contains ξ_1 but not ξ_2, and C_{d_2} contains ξ_2 but not ξ_1. Because of Theorems 3.9 and 3.7, C_{d_1} and C_{d_2} are closed and convex. Furthermore, we obviously have

$$g_2W_{21} \leqq g_1W_{12} \text{ for all } \xi \text{ in } C_{d_1} \tag{4.2}$$

and

$$g_2W_{21} \geqq g_1W_{12} \text{ for all } \xi \text{ in } C_{d_2}. \tag{4.3}$$

Let $\xi_0 = (g_1^0, g_2^0)$ be the a priori distribution for which

$$g_2^0W_{21} = g_1^0W_{12}. \tag{4.4}$$

It follows from (4.2) and (4.3) that there exist two positive numbers c' and c'' such that

$$0 < c' \leqq g_2^0 \leqq c'' < 1 \tag{4.5}$$

and such that the class C_{d_1} consists of all ξ for which $g_2 \leqq c'$, and the class C_{d_2} consists of all ξ for which $g_2 \geqq c''$.

Thus, the following decision procedure will be a Bayes solution relative to the a priori distribution $\xi = (g_1, g_2)$: *If $g_2 \leqq c'$ or $\geqq c''$, do not take any observations and make the corresponding final decision. If $c' < g_2 < c''$, continue taking observations until the a posteriori probability of H_2 is either $\geqq c''$ or $\leqq c'$. If this a posteriori probability is $\geqq c''$, accept H_2, and if it is $\leqq c'$, accept H_1.*

The a posteriori probability of H_2 after the first m observations have been made is given by

$$(4.6) \qquad g_{2m} = \frac{g_2 p(x_1 \mid F_2) \cdots p(x_m \mid F_2)}{g_1 p(x_1 \mid F_1) \cdots p(x_m \mid F_1) + g_2 p(x_1 \mid F_2) \cdots p(x_m \mid F_2)}.$$

If $c' < g_2 < c''$ and if the probability (under F_1 as well as under F_2) is zero that $g_{2m} = c'$ or $= c''$ for some m, then it follows from Theorem 3.8 that the above described Bayes solution is essentially unique; i.e., any other Bayes solution can differ from the one given above only on a set whose probability measure is zero under both F_1 and F_2.

Provided that at least one observation is made, one can easily verify that the above described Bayes solution is identical with a sequential probability ratio test for testing H_2 against H_1. The sequential probability ratio test is defined as follows (see [3]): Two positive constants A and B ($B < A$) are chosen. Experimentation is continued as long as the probability ratio

$$(4.7) \qquad \frac{p_{2m}}{p_{1m}} = \frac{p(x_1 \mid F_2) \cdots p(x_m \mid F_2)}{p(x_1 \mid F_1) \cdots p(x_m \mid F_1)}$$

satisfies the inequality $B < \frac{p_{2m}}{p_{1m}} < A$. If $\frac{p_{2m}}{p_{1m}} \geqq A$, accept H_2. If $\frac{p_{2m}}{p_{1m}} \leqq B$, accept H_1. The Bayes solution described above coincides with this probability ratio test for properly chosen values of the constants A and B

The results described above for $k = 2$ are essentially the same as those contained in Lemmas 1 and 2 of an earlier publication [4] of the authors

We shall now discuss the case $k = 3$. Any a priori distribution ξ can be represented by a point with the barycentric coordinates g_1, g_2 and g_3, where g_i is the a priori probability of $H_i (i = 1, 2, 3)$. The totality of all possible a priori distributions ξ will fill out the triangle T with the vertices 0_1, 0_2 and 0_3 where 0_i represents the a priori distribution corresponding to $g_i = 1$ (see Figure 1).

Clearly, the vertex 0_i is contained in C_{d_i}. Thus, because of Theorem 3.9, $C_{d_i} (i = 1, 2, 3)$ is a convex subset of T containing the vertex 0_i, as indicated in Figure 1.

If one of the components of $\xi = (g_1, g_2, g_3)$ is zero, say $g_i = 0$, then H_i can be disregarded and the problem of constructing Bayes solutions reduces to the previously considered case where $k = 2$. Thus, in particular, the determination of the boundary points $P_1, P_2, \cdots, P_6$ of C_{d_1}, C_{d_2} and C_{d_3} which are on the boundary of the triangle T, reduces to the previously considered case, $k = 2$.

It follows from Theorems 3.8 and 3.9 that the intersection of C_{d_t} with any straight line T_t through 0_t is a closed segment. One endpoint of this segment is, of course, 0_t. Let B_t denote the other endpoint. It follows from Theorem 3.7 that B_t must be a point of C_{2,d_t}. Any interior point of 0_tB_t can be shown to be an element of C_{1,d_t}. The proof of this is very similar to that of Theorem 3.9

We shall now show how tangents to the sets C_{d_1}, C_{d_2} and C_{d_3} can be constructed at the boundary points $P_1, P_2, \cdots, P_6$. Consider, for example, the boundary point P_1 of C_{d_1} that lies on the line $0_1\, 0_2$. Let ξ_1 be the a priori distribution represented by the point P_1. Since the a priori probability of H_3 is zero according to ξ_1, we can disregard H_3 in constructing Bayes solutions relative to ξ_1. Let $D_1(x)$ be a sequential probability ratio test for testing H_1 against H_2

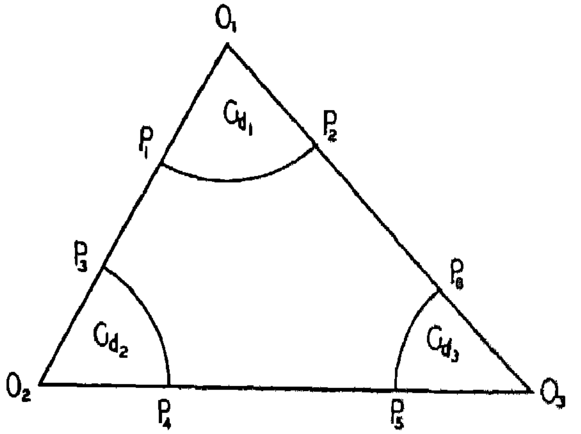


FIG. 1

which requires at least one observation and which is a Bayes solution relative to ξ_1. Since ξ_1 is a boundary point, such a decision function D_1 exists. Thus, we have

$$W(\xi_1, d_1) = r(\xi_1, D_1) = \operatorname*{Inf}_{D} r(\xi_1, D). \tag{4.8}$$

Let α_{ij} denote the probability of accepting H_j when H_i is true and D_1 is the decision function adopted. Let, furthermore, n_i denote the expected number of observations required by the decision procedure when F_i is true and D_1 is adopted Then, for any a priori distribution $\xi = (g_1, g_2, g_3)$ we have

$$r(\xi, D_1) = \sum_{i,j} g_i W_{ij} \alpha_{ij} + \sum_i g_i n_i \tag{4.9}$$

and

$$W(\xi, d_1) = \sum_i g_i W_{i1}. \tag{4.10}$$

Thus, the linear manifold $L(D_1, d_1)$ is simply the straight line given by the equation

$$\sum_i g_i W_{i1} = \sum_{i,j} g_i W_{ij} \alpha_{ij} + \sum_i g_i n_i . \tag{4.11}$$

This straight line goes through P_1 and, because of Theorem 3.10, it is tangent to C_{d_1}. Tangents at the same points $P_2, \cdots, P_6$ can be constructed in a similar way.

The convexity properties of the sets $C_{d_i}(i = 1, 2, \cdots, k)$ were established by the authors prior to the more general results described in Sections 2 and 3 and were stated by one of the authors in an address given at the Berkeley meeting of the Institute of Mathematical Statistics, June, 1948. More general results when Ω and D^* are finite, admitting also non-linear cost functions, were obtained later by Arrow, Blackwell and Girshick [2].

REFERENCES

[1] A. WALD, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (1947), pp. 279–313.

[2] K. J. ARROW, D. BLACKWELL, M. A. GIRSHICK, "Bayes and minimax solutions of sequential decision problems," *Econometrica*, Vol. 17 (1949), pp. 213–244.

[3] A. WALD, "*Sequential analysis*," John Wiley & Sons, New York, 1947.

[4] A. WALD AND J. WOLFOWITZ, "Optimum character of the sequential probability ratio test," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 326–339.

[5] S. SAKS, "*Theory of the integral*," Hafner Publishing Company, New York.

[6] A. WALD, "Statistical decision functions," *Annals of Math. Stat*, Vol. 20 (1949), pp 165–205.

ON THE DISTRIBUTIONS OF MIDRANGE AND SEMI-RANGE IN SAMPLES FROM A NORMAL POPULATION

K. C. S. Pillai

University of Travancore, Trivandrum.

1. Summary. In this paper the simultaneous distribution of midrange and semi-range has been obtained and used to derive the distributions of midrange and semi-range in samples taken from a normal population.

2. Introduction. The concept of ordering a sample has given rise to innumerable problems for statistical investigation. Several authors have contributed to the study of ordered individuals and, in particular, to the study of extreme individuals, their sum and difference in samples from a normal population. L. H. C. Tippett [1] has studied the first four moments of the range and has tabled the mean-range for sample size ranging from two to thousand Student [2] has determined the nature of the distribution of range for particular sample sizes by purely empirical methods. T. Hojo [3] has compared the standard error of midrange to that of median and mean in normal samples. E. S. Pearson and H. O Hartley [4] have tabled the values of the probability integral of range for sample size up to twenty. E. J. Gumbel [5], [6], [7] has established the independence of the extreme values in large samples from population of unlimited range and obtained the distributions of range and midrange. The asymptotic distribution of range has also been investigated by G. Elfving [8]. J. F. Daly [9] has devised a t-test adopting range in place of standard deviation in Student's t and in a modified t-test E Lord [10] has used range instead of standard deviation. An extension to two populations of an analogue of Student's t-test using the sample range has been worked out by John E. Walsh [11]. S. S. Wilks [12] has given a complete and detailed account of the researches on order statistics and also a number of suggestions regarding possibilities of utilising order statistics in statistical inference. In this paper the distribution of midrange has been developed as a series and a method of evaluating the probability integral for semi-range based on an infinite series expansion for the normal probability integral has been suggested.

3. Distributions of midrange and semi-range. Let

$$x_1 \leq x_2 \cdots \leq x_n$$

be an ordered sample from a normal population with zero mean and unit standard deviation. Then the joint distribution of x_1 and x_n, the lowest and highest values respectively, is given by [13],

$$(1)\qquad p(x_1, x_n) = [n(n-1)/2\pi]\left[\int_{x_1}^{x_n} e^{-t^2/2}\,dt/\sqrt{2\pi}\right]^{n-2} e^{-(x_1^2+x_n^2)/2}.$$

Let

$$M = (x_1 + x_n)/2$$

and

$$W = (x_n - x_1)/2.$$

M is the midrange and W is the semi-range of the sample. From (1) the simultaneous distribution of M and W reduces to

$$(2)\qquad p(M, W) = [n(n-1)/\pi]e^{-(M^2+W^2)}\left[\int_{M-W}^{M+W} e^{-t^2/2}\,dt/\sqrt{2\pi}\right]^{n-2}.$$

It has been shown [14] that if

$$(3)\qquad F(M, W) = \left[\int_{M-W}^{M+W} e^{-t^2/2}\,dt/\sqrt{2\pi}\right]^{k},$$

$$(4)\qquad F(M, W) = e^{-k(M^2+W^2)/2}[A_0^{(k)} + A_1^{(k)}M^2 + \cdots + A_i^{(k)}M^{2i} + \cdots],$$

where $A_i^{(k)}$ coefficient is given by

$$(5)\qquad \begin{aligned} 2iA_i^{(k)} = kA_{i-1}^{(k)} - k\sqrt{2/\pi}[A_{i-1}^{(k-1)}W \\ + A_{i-2}^{(k-1)}W^3/\Gamma(4) + \cdots + A_0^{(k-1)}W^{2i-1}/\Gamma(2i)]. \end{aligned}$$

Using expansion (4) equation (2) reduces to

$$(6)\qquad p(M, W) = [n(n-1)/\pi]e^{-n(M^2+W^2)/2}\sum_{i=0}^{\infty} A_i^{(n-2)} M^{2i}.$$

It is evident that the A's involve terms of the form

$$[\phi(W)]^s W^q e^{-mW^2/2}$$

where s, q, m are positive integers and

$$\phi(W) = \sqrt{2/\pi}\int_0^W e^{-t^2/2}\,dt.$$

Integrating (6) with respect to W

$$(7)\qquad p(M) = [n(n-1)/\pi]e^{-nM^2/2}\sum_{i=0}^{\infty} B_i M^{2i}$$

where

$$(8)\qquad B_0 = \sqrt{\pi/2}\,I(n-2, 0, 2),$$

$$(9)\qquad \begin{aligned} B_1 &= [(n-2)/2][\sqrt{\pi/2}\,I(n-2, 0, 2) - I(n-3, 1, 3)], \\ B_2 &= [(n-2)/2^2\Gamma(3)][\sqrt{\pi/2}\,(n-2)I(n-2, 0, 2) \end{aligned}$$

$$(10) \qquad - (2n - 5)I(n - 3, 1, 3) - (1/3)I(n - 3, 3, 3) + \sqrt{2/\pi}\,(n - 3)I(n - 4, 2, 4)]$$

where

$$(11) \qquad I(s, q, m) = \sqrt{2/\pi} \int_0^\infty [\phi(x)]^s x^q e^{-mx^2/2}\, dx.$$

Using the method of integration by parts, the evaluation of $I(s, q, m)$ can be reduced ultimately to that of $I(p, 0, r)$ and this function for different values of p and r is given in Table I.

TABLE I

Values of Integrals $I(p, 0, r)$[1]

p	r			
	2	4	6	8
1	0.277,063,21	0.147,583,62	0.100,735,97	0.076,490,19
2	0.152,980,4	0.064,094,20	0.037,255,93	0.025,060,53
3	0.098,373	0.033,453,6	0.016,808,71	
4	0.069,10	0.019,535,1	0.008,589,57	
5	0 051,44	0.012,325,5		
6	0 039,90	0 008,223,9		
7	0 031,94			
8	0 026,17			

The first five B Coefficients for n ranging from 3 to 10 are tabled below.

TABLE II

Values of B Coefficients.

n	B_0	B_1	B_2	B_3	B_4
3	0 347,247,25	0 040,642,87	0.002,772,90	0.000,133,80	0.000,005,00
4	0 191,732	0 058,751	0 010,906	0.001,460	0.000,153
5	0.123,292	0.067,184	0.021,526	0.004,988	0.000,909
6	0.086,60	0.070,93	0.033,23	0.011,20	0.002,97
7	0 064,47	0.072,20	0.045,65	0.020,28	0.007,14
8	0.050,01	0.072,09	0.057,22	0.032,21	0.014,59
9	0.040,03	0.071,27	0.068,95	0.047,01	0.024,98
10	0 032,80	0.069,97	0.080,31	0.064,66	0.040,51

[1] The integrals have been evaluated by using (14).

The accuracy obtained by keeping the first five terms in $p(M)$ may be judged from the following values of the total probability calculated for small values of n.

TABLE III.

Total probability keeping the first five terms in $p(M)$

Size of sample	3	4	5	6	7
Total probability	0.999,998	0.999,92	0.999,56	0 998,8	0.997,8

Integrating (6) with respect to M, $p(W)$ may be obtained But $p(W)$ involves integral $\phi(W)$ and to evaluate the integral probability of W expansions for $\phi(W)$ and its powers have to be developed.

Since $$\phi(W) = \sqrt{2/\pi}\int_0^W e^{-t^2/2}\,dt = \sqrt{2/\pi}\,W(1 - W^2/6 + \cdots),$$

a convenient expansion is given by

$$\sqrt{2/\pi}\int_0^W e^{-t^2/2}\,dt = \sqrt{2/\pi}\,We^{-W^2/6}(1 + a_2W^4 + \cdots + a_iW^{2i} + \cdots) \tag{12}$$

where a_i follows the recurrence relation

$$3(2i + 1)a_i - a_{i-1} = (-1)^i/3^{i-1}\Gamma(i + 1), \tag{13}$$

as may be seen by differentiating (12) with respect to W and equating the coefficient of W^{2i} on both sides. Again

$$[\phi(W)]^j = (2/\pi)^{j/2}e^{-jW^2/6}W^jS^j \tag{14}$$

where

$$S = 1 + a_2W^4 + a_3W^6 + \cdots + a_iW^{2i} + \cdots \tag{15}$$

and

$$S^j = 1 + K_2^{(j)}W^4 + K_3^{(j)}W^6 + \cdots \tag{16}$$

where

$$K_i^{(j)} = \sum_{s=1}^{i} {}^jC_s\, s!\, a_1^{s_1}a_2^{s_2}\cdots a_i^{s_i}/s_1!\,s_2!\cdots s_i! \tag{17}$$

and

$$\begin{aligned} s_1 + 2s_2 + \cdots + is_i &= i, \\ s_1 + s_2 + \cdots + s_i &= s. \end{aligned} \tag{17a}$$

Clearly $a_i = K_i^{(1)}$. In evaluating the $K_i^{(j)}$'s summation with respect to s is first

performed, the values of $s_1, s_2, \cdots, s_i$ being obtained so as to satisfy the relations (17a); and thereafter the values of the a's are substituted. It may be noted that $a_1 = 0$ The K coefficients for j up to 8 and i up to 13 are given below.

TABLE IV

$K_{ij}^{(i)}$ Coefficients.

j	i			
	2	3	4	5
1	0.011,111,11	$-0.0^{3}35,273,369$	$0.0^{4}44,091,711$	$-0.0^{5}17,814,833$
2	0.022,222,22	$-0.0^{3}70,546,737$	$0.0^{3}21,164,021$	$-0.0^{4}11,401,493$
3	0.033,333,33	$-0.0^{2}10,582,011$	$0.0^{3}50,264,550$	$-0.0^{4}28,860,029$
4	0.044,444,44	$-0.0^{2}14,109,348$	$0.0^{3}91,710,758$	$-0.0^{4}54,157,091$
5	0.055,555,56	$-0.0^{2}17,636,684$	$0.0^{2}14,550,265$	$-0.0^{4}87,292,680$
6	0.066,666,67	$-0.0^{2}21,164,021$	$0.0^{2}21,164,021$	$-0.0^{3}12,826,680$
7	0.077,777,78	$-0.0^{2}24,691,358$	$0.0^{2}29,012,346$	$-0.0^{3}17,707,944$
8	0.088,888,89	$-0.0^{2}28,218,695$	$0.0^{2}38,095,238$	$-0.0^{3}23,373,061$

j	i			
	6	7	8	9
1	$0.0^{6}10,087,459$	$-0.0^{8}38,065,882$	$0.0^{9}14,772,299$	$-0.0^{11}47,770,889$
2	$0.0^{5}13,059,860$	$-0.0^{7}78,306,957$	$0.0^{8}57,379,607$	$-0.0^{9}32,240,604$
3	$0.0^{5}49,870,764$	$-0.0^{6}35,414,321$	$0.0^{7}37,246,865$	$-0.0^{8}26,934,251$
4	$0.0^{4}12,515,888$	$-0.0^{6}96,195,746$	$0.0^{6}13,039,809$	$-0.0^{7}10,793,811$
5	$0.0^{4}25,264,163$	$-0.0^{5}20,323,918$	$0.0^{6}33,614,797$	$-0.0^{7}30,234,979$
6	$0.0^{4}44,603,642$	$-0.0^{5}36,960,883$	$0.0^{6}72,070,037$	$-0.0^{7}68,563,784$
7	$0.0^{4}71,905,926$	$-0.0^{5}60,836,892$	$0.0^{5}13,654,992$	$-0.0^{6}13,526,252$
8	$0.0^{3}10,854,319$	$-0.0^{5}93,258,365$	$0.0^{5}23,672,301$	$-0.0^{6}24,174,801$

j	i			
	10	11	12	13
1	$0.0^{12}14,640,444$	$-0.0^{14}40,268,872$	$0.0^{15}10,359,029$	$-0.0^{17}24,535,539$
2	$0.0^{10}18,330,114$	$-0.0^{12}91,351,579$	$0.0^{13}43,595,840$	$-0.0^{14}19,132,452$
3	$0.0^{9}21,506,514$	$-0.0^{10}14,469,203$	$0.0^{12}96,661,910$	$-0.0^{13}58,727,628$
4	$0.0^{8}10,849,591$	$-0.0^{10}87,178,260$	$0.0^{11}72,767,557$	$-0.0^{12}54,213,617$
5	$0.0^{8}36,260,639$	$-0.0^{9}32,719,538$	$0.0^{10}32,219,900$	$-0.0^{11}27,049,719$
6	$0.0^{8}95,092,297$	$-0.0^{9}93,120,388$	$0.0^{9}10,472,881$	$-0.0^{11}96,020,717$
7	$0.0^{7}21,247,442$	$-0.0^{8}22,112,968$	$0.0^{9}27,825,332$	$-0.0^{10}27,369,553$
8	$0.0^{7}42,365,199$	$-0.0^{8}46,218,579$	$0.0^{9}64,147,144$	$-0.0^{10}66,862,484$

Using (12) the probability integral for W can be evaluated with the help of tables of Incomplete Gamma Functions.

REFERENCES

[1] L H C. TIPPET, "On the extreme individuals and the range of samples taken from a normal population," *Biometrika*, Vol 17 (1925), pp 364–387.

[2] STUDENT, "Errors in routine analysis," *Biometrika*, Vol 19 (1927), pp. 151–164.

[3] T HOJO, "Distribution of the median, quartiles and interquartile distance in samples from a normal population," *Biometrika*, Vol. 23 (1931), pp 315–360.

[4] E. S. PEARSON AND H. O HARTLEY, "The probability integral of the range in samples of n observations from a normal population," *Biometrika*, Vol 32 (1942), pp. pp. 301–310.

[5] E. J. GUMBEL, "Ranges and midranges," *The Annals of Math. Stat.*, Vol. 15 (1944), No. 4, pp 414–422

[6] E. J. GUMBEL, "On the independence of the extremes in a sample," *The Annals of Math. Stat*, Vol. 17 (1946), No. 1, pp. 78–80.

[7] E J. GUMBEL, "The distribution of the range," *The Annals of Math. Stat.*, Vol. 18 (1947), No. 3, pp. 384–412.

[8] G. ELFVING, "The asymptotical distribution of range in samples from a normal population," *Biometrika*, Vol. 34, (1947), pp 111–119

[9] J. F. DALY, "On the use of the sample range in an analogue of Student's t-test," *The Annals of Math. Stat.*, Vol 17 (1946), No. 1, pp. 71–74

[10] E LORD, "The use of range in place of standard deviation in the t-test," *Biometrika*, Vol 34 (1947), pp. 41–67.

[11] J E WALSH, "An extension to two populations of an analogue of Student's t-test using the sample range," *Annals of Math. Stat.*, Vol. 18 (1947), No. 2, pp. 280–285.

[12] S S. WILKS, "Order Statistics," *Am Math. Soc Bull*, Vol 54 (1948), pp. 6–50.

[13] S. S WILKS, *Mathematical Statistics*, Princeton University Press, Princeton, 1943, p. 90.

[14] K C. S. PILLAI, "A note on ordered samples," *Sankhya*, Vol. 8 (1948), Part 4, pp. 375–380.

THE IMPOSSIBILITY OF CERTAIN SYMMETRICAL BALANCED INCOMPLETE BLOCK DESIGNS

By S. S. Shrikhande

University of North Carolina

Introduction and Summary. An arrangement of v varieties or treatments in b blocks of size k, $(k < v)$, is known as a balanced incomplete block design if every variety occurs in r blocks and any two varieties occur together in λ blocks. These parameters obviously satisfy the equations

$$bk = vr \tag{1}$$

$$\lambda(v - 1) = r(k - 1). \tag{2}$$

Fisher [1] has also proved that the inequality

$$b \geq v, \qquad r \geq k \tag{3}$$

must hold. If v, b, r, k and λ are positive integers satisfying (1), (2) and (3), then a balanced incomplete block design with these parameters possibly exists, but the actual existence of a combinatorial solution is not ensured. These conditions are thus necessary but not sufficient for the existence of a design. Fisher and Yates in their tables [2] have listed all designs with $r \leq 10$ and given combinatorial solutions, where known. A balanced incomplete block design in which $b = v$, and hence $r = k$ is called a symmetrical balanced incomplete block design. The impossibility of the symmetrical designs with parameters $v = b = 22$, $r = k = 7$, $\lambda = 2$ and $v = b = 29$, $r = k = 8$, $\lambda = 2$ was first demonstrated by Hussain [3], [4] essentially by the method of enumeration. The object of the present note is to give an alternative simple proof of the impossibility of these designs and to show that the only unknown remaining symmetrical design in Fisher and Yates' tables, *viz.* $v = b = 46$, $r = k = 10$, $\lambda = 2$, is definitely impossible. Symmetrical designs with $\lambda \leq 5$, r, $k \leq 20$, which are impossible combinatorially, are also listed.

1. A necessary condition for the existence of a symmetrical balanced incomplete block design when v is even.

Theorem 1. *A necessary condition for the existence of a symmetrical balanced incomplete block design with parameters v, r and λ, where v is even, is that $r - \lambda$ be a perfect square.*

Proof. Let $N = (n_{ij})$ be a square matrix of v rows and v columns where

$$n_{ij} = 1 \text{ or } 0 \tag{4}$$

according as the i-th treatment does or does not occur in the j-th block. Put

$$B = NN' \tag{5}$$

Since every treatment occurs in r blocks and every pair of treatments in λ blocks, we have, if the design is possible,

$$B = \begin{pmatrix} r & \lambda & \cdots & \lambda \\ \lambda & r & \cdot\ \cdot & \lambda \\ \cdots & & \cdots & \cdot \\ \lambda & \lambda & \cdots & r \end{pmatrix}. \tag{6}$$

Subtracting the first column from all the other columns and then adding to the first row all the other rows, we see that

$$\begin{aligned} |B| &= [r + \lambda(v-1)](r-\lambda)^{v-1} \\ &= r^2(r-\lambda)^{v-1} \text{ from (2).} \end{aligned} \tag{7}$$

But from (5)

$$|B| = |N|^2.$$

Since $|N|$ is integral, it follows that $(r-\lambda)^{v-1}$ is the square of an integer, and hence if v is even, $r - \lambda$ must be a perfect square.

COROLLARY. *The following symmetrical designs are impossible.*

(A_1)	$v = b = 22$	$r = k = 7$	$\lambda = 2$
(A_2)	$v = b = 46$	$r = k = 10$	$\lambda = 2$
(A_3)	$v = b = 92$	$r = k = 14$	$\lambda = 2$
(A_4)	$v = b = 106$	$r = k = 15$	$\lambda = 2$
(A_5)	$v = b = 172$	$r = k = 19$	$\lambda = 2$
(A_6)	$v = b = 34$	$r = k = 12$	$\lambda = 4.$

As already mentioned in the introduction, the impossibility of (A_1) has been proved by Hussain [3], but for the design (A_2) it was hitherto unknown whether or not a solution is possible and it was left as a blank in the latest edition of Fisher and Yates' tables.

2. Application of method of Bruck and Ryser.

In a recent paper Bruck and Ryser [5] have proved the impossibility of some finite projective planes with the help of the properties of matrices whose elements are integers. Their method is immediately applicable to our own problem.

Let A and B be two symmetric matrices of order n with elements in the rational field. The matrices A and B are congruent, written $A \sim B$, provided there exists a nonsingular matric C with elements in the rational field, such that $A = C'BC$. The congruence of matrices satisfies the usual requirements of an "equals" relationship.

If A is an integral symmetric matrix of order n and rank n, we can always construct an integral diagonal matrix $D = (d_1, \cdots, d_n)$, where $d_i \neq 0$, $i = 1, 2, \cdots n$ such that $D \sim A$ The number of negative terms i, called the index of A, is an invariant by Sylvester's Law.

Define $d = (-1)^s\delta$ where δ is the square-free positive part of $|A|$. Then since $|B| = |C|^2|A|$, d is another invariant of A

Now let A be a nonsingular and symmetric integral matrix of order n. Let D_r be the leading principal minor determinant of order r and suppose that $D_r \neq 0$ for $r = 1, 2, \cdots n$. Define

$$C_p(A) = (-1, -D_n)_p \prod_{j=1}^{n-1} (D_j, -D_{j+1})_p \tag{9}$$

for every odd prime p where $(m, m')_p$ is the Hilbert norm-residue symbol for arbitrary non-zero integers m and m' and for every prime p. The following two theorems are given in the collected works of Hilbert [6].

THEOREM (A) *If m and m' are integers not divisible by the odd prime p, then*

$$(m, m')_p = +1 \tag{10}$$

$$(m, p)_p = (p, m)_p = (m/p), \tag{11}$$

where (m/p) is the Legendre symbol. Moreover, if $m \equiv m' \not\equiv 0 \bmod p$, then

$$(m, p)_p = (m', p)_p . \tag{12}$$

THEOREM (B). *For arbitrary non-zero integers m, m', n, n' and for every prime p,*

$$(-m, m)_p = +1 \tag{13}$$

$$(m, n)_p = (n, m)_p \tag{14}$$

$$(mm', n)_p = (m, n)_p(m', n)_p \tag{15}$$

$$(m, nn')_p = (m, n)_p(m, n')_p . \tag{16}$$

From the above it is easy to prove that for p an odd prime and every positive integer m,

$$(m, m+1)_p = (-1, m+1)_p \tag{17}$$

$$\prod_{j=1}^{m} (j, j+1)_p = ((m+1)!, -1)_p . \tag{18}$$

We can now state the fundamental Minkowski-Hasse Theorem [7].

THEOREM (C). *Let A and B be two integral symmetric matrices of order n and rank n. Suppose further that the leading principal minor determinants of A and B are different from zero. Then $A \sim B$ if and only if A and B have the same invariants i, d and C_p for every odd prime p.*

3. A necessary condition for the existence of a symmetrical balanced incomplete block design for any integer v.

Suppose the symmetrical design with parameters v, r and λ exists. Then with the previous definition of N and B,

$$B = NN' = \begin{pmatrix} r & \lambda & \cdots & \lambda \\ \lambda & r & \cdots & \lambda \\ \cdots & & \cdots & \\ \lambda & \lambda & \cdots & r \end{pmatrix}.$$

Subtracting the last row from the remaining rows and then subtracting the last column from all the other columns, we get

$$(19)\qquad Q = \begin{bmatrix} 2(r-\lambda) & (r-\lambda) & \cdots & (r-\lambda) & -(r-\lambda) \\ (r-\lambda) & 2(r-\lambda) & \cdots & (r-\lambda) & -(r-\lambda) \\ \cdots & & \cdots & & \cdots \\ (r-\lambda) & (r-\lambda) & \cdots & 2(r-\lambda) & -(r-\lambda) \\ -(r-\lambda) & -(r-\lambda) & \cdots & -(r-\lambda) & r \end{bmatrix}.$$

Obviously $Q \sim B$. But $B \sim I$. Hence $Q \sim I$ and, therefore, since Q and I satisfy all the conditions of Theorem C, they must have the same invariants i, d and C_p.

Let D_j denote the leading principal minor determinant of Q of order j. Then

$$(20)\qquad D_j = (r-\lambda)^j(j+1) \qquad \text{for } j = 1, 2, \cdots, v-1$$

$$(21)\quad \text{and} \qquad D_v = |B| = r^2(r-\lambda)^{v-1}.$$

Then, omitting p for convenience,

$$C_p(Q) = (-1, -D_v)(D_{v-1}, -D_v)\prod_{j=1}^{v-2}(D_j, -D_{j+1}).$$

We use (10) $\cdot$, (18) in deriving the value of $C_p(Q)$.

Now

$$\begin{aligned} (-1, -D_v)(D_{v-1}, -D_v) & \\ &= (-1, -r^2(r-\lambda)^{v-1})((r-\lambda)^{v-1}v, -r^2(r-\lambda)^{v-1}) \\ &= (-1,-1)(-1,r^2)(-1,(r-\lambda)^{v-1})((r-\lambda)^{v-1}, r^2) \\ &\quad ((r-\lambda)^{v-1}, -(r-\lambda)^{v-1})(v, r^2)(v, -(r-\lambda)^{v-1}) \\ &= (-1, (r-\lambda)^{v-1})(v, -(r-\lambda)^{v-1}) \\ &= (-1, (r-\lambda)^{v-1})(v, -1)(v, (r-\lambda))^{v-1}. \end{aligned}$$

Also

$$\prod_{j=1}^{v-2}(D_j, -D_{j+1}) = \prod_1^{v-2}((r-\lambda)^j(j+1), -(r-\lambda)^{j+1}(j+2))$$

$$\begin{aligned} &= \left\{\prod_1^{v-2}((r-\lambda)^j, -(r-\lambda)^{j+1})(j+1, -(j+2))\right\} S \\ &= S\prod_1^{v-2}((r-\lambda)^j, -(r-\lambda)^j)((r-\lambda)^j, (r-\lambda))(j+1, j+2)(j+1, -1) \\ &= S\prod_1^{v-2}((r-\lambda), (r-\lambda))^j(j+2, -1)(j+1, -1) \\ &= S\prod_1^{v-2}(r-\lambda, -1)^j(j+2, -1)(j+1, -1) \\ &= S(r-\lambda, -1)^{(v-1)(v-2)/2}((v-1)!, -1)(v!, -1) \\ &= S(r-\lambda, -1)^{(v-1)(v-2)/2}(v, -1), \end{aligned}$$

where

$$
\begin{aligned}
S &= \prod_{1}^{v-2} ((r-\lambda)^j, j+2)((r-\lambda)^{j+1}, j+1) \\
&= \prod_{1}^{v-2} ((r-\lambda)^j, j+2)((r-\lambda)^{j-1}, j+1) \\
&= \prod_{1}^{v-2} ((r-\lambda)^j, j+2) \prod_{j=0}^{v-3}((r-\lambda)^j, j+2) \\
&= (r-\lambda, v)^{v-2}. \\
\therefore\ C_p(Q) &= (r-\lambda, -1)^{v(v-1)/2} (v, -1)^2 (r-\lambda, v)^{2v-3}
\end{aligned}
$$

(22) $$= (r-\lambda, -1)^{v(v-1)/2}(r-\lambda, v)^{2v-3}.$$

Hence we can enunciate the following theorem:

THEOREM 2. *A necessary condition for the existence of a symmetrical balanced incomplete block design with parameters v, r and λ is that*

$$C_p(Q) = (r-\lambda, -1)_p^{v(v-1)/2}(r-\lambda, v)_p^{2v-3} = +1$$

for all odd prime p, where $(m, n)_p$ is the Hilbert norm-residue symbol.

When v is even we have seen that a necessary condition for the existence of the design is that $r - \lambda$ be a perfect square. Then it is easily seen that

$$C_p(Q) = +1$$

for all odd prime p. Therefore, even if the design is really non-existent, its impossibility cannot be proved by this method.

When, however, v is odd we can in many instances demonstrate the impossibility of the design.

Consider the design

(A_7) $$v = b = 29, \qquad r = k = 8, \qquad \lambda = 2.$$

$$
\begin{aligned}
C_p(Q) &= (6, -1)_p^{29\cdot 14}(6, 29)_p^{55} \\
&= (3, 29)_p(2, 29)_p \\
&= (29/3) \text{ for p} = 3 \\
&= (2/3) \text{ for p} = 3 \\
&= -1 \quad \text{for p} = 3.
\end{aligned}
$$

Hence the design (A_7) is impossible. As mentioned in the introduction, the impossibility has already been demonstrated by Hussain [4] by a rather lengthy method amounting to a complete exhaustion of all possibilities. The following designs with $\lambda \leq 5$ and $r, k \leq 20$ can be similarly proved to be impossible by applying Theorem 2.

(A_8)	$v = b = 137$	$r = k = 17$	$\lambda = 2$
(A_9)	$v = b = 67$	$r = k = 12$	$\lambda = 2$
(A_{10})	$v = b = 103$	$r = k = 18$	$\lambda = 3$
(A_{11})	$v = b = 53$	$r = k = 13$	$\lambda = 3$
(A_{12})	$v = b = 43$	$r = k = 15$	$\lambda = 5$
(A_{13})	$v = b = 77$	$r = k = 20$	$\lambda = 5.$

My thanks are due to Professor R. C. Bose under whose guidance this research was carried out.

REFERENCES

[1] R. A. FISHER, "An examination of the different possible solutions of a problem in incomplete blocks," *Annals of Eugenics*, Vol 10 (1940), pp. 52–75.

[2] R. A. FISHER AND F. YATES, *Statistical Tables for Biological Agricultural and Medical Research* (1949), Hafner Publishing Company, New York.

[3] Q. M. HUSSAIN, "Impossibility of the symmetrical incomplete block design with $\lambda = 2$, $k = 7$," *Sankhyā*, Vol. 7 (1946), pp. 317–322.

[4] Q. M. HUSSAIN, "Symmetrical incomplete block design with $\lambda = 2$, $k = 8$ or 9," *Bulletin of the Calcutta Mathematical Society*, XXXVII (1945), pp. 115–123.

[5] R. H. BRUCK AND H. J. RYSER, "Non-existence of certain finite projective planes," *Canadian Jour. Math.*, Vol. 1 (1949), pp. 88–93.

[6] D. HILBERT, *Gesammelte Abhandlungen* I (Berlin, 1932), pp. 161–173.

[7] H. HASSE, "Über die Äquivalenz quadratischer formen im korper rationalen zahlen," *J. Reine Angew. Math.*, Vol 152 (1923), pp. 205–224.

NOTES

This section is devoted to brief research and expository articles and other short items.

THE SAMPLING DISTRIBUTION OF THE RATIO OF TWO RANGES FROM INDEPENDENT SAMPLES[1]

RICHARD F. LINK
University of Oregon

Let us consider a sample of n ordered observations $(x_1 < x_2 < \cdots < x_n)$ drawn from a population with variance σ^2. Let $w = (x_n - x_1)/\sigma$. Let us consider the joint sampling distribution of w_1 and w_2 for two samples, not necessarily the same size, drawn from populations with the same variance. If the two samples were drawn independently, then the joint sampling distributions of w_1 and w_2 may be written as the product of the sampling distributions of w_1 and w_2.

If we make the change of variable $r = w_1/w_2$, $w_2 = w$, and if w is integrated over its range of definition, the cumulative distribution of the ratio of two ranges remains. This may be written as

$$F(R) = \int_0^R dr \int_0^\infty dw \cdot w \cdot h_2(w) \cdot h_1(wr), \tag{1}$$

where h_1 is the pdf for w_1 and h_2 is the pdf for w_2.

To obtain more explicit results, specific distribution functions may be considered. The following table gives the sampling distribution of the ratio of two ranges from independent samples for the indicated density functions $f(x)$. Notice that for the normal distribution it was possible to obtain results only for some special cases.

In Table 1 for $F(R)$, w_1 and w_2 represent ranges computed from samples of size n_1 and n_2 respectively.

Notice that formula (1) for $F(R)$ is equivalent to the following expressions

$$Pr(w_1/w_2 < R) = F(R) = \int_0^\infty dw_2 \int_0^{Rw_2} dw_1\, h(w_1) \cdot h(w_2).$$

The region of integration for the last expression is simply the region in the w_2, w_1 plane to the right of the line $w_1 = Rw_2$.

This integration was done numerically. Table 2 gives values of R for all combinations of n_1 and $n_2 \leq 10$ and for $\alpha = .005, .01, .025, .05, .10$ such that

$$Pr(w_1/w_2 < R) = \alpha$$

where w_1 and w_2 are ranges computed from samples of size n_1 and n_2 drawn from

[1] This work was done under contract N6onr-218/IV with the Office of Naval Research.

normal populations with the same variance. It is believed that these values are correct to within one place in the last reported figure.

These tabled values may be used as critical values for testing the hypothesis that two independent samples were drawn from normal populations with the same variance. This test is therefore comparable to the F test. Some sort of

TABLE 1

$f(x)$	$F(R) = Pr\ (w_1/w_2 < R)$
$1 \quad 0 \leq x \leq 1$ $0 \quad$ all other x	$n_2(n_2-1)\ R^{n_1-1}\left[\frac{n_1}{n_1+n_2-2} - \frac{(n_1+R[n_1-1])}{n_1+n_2-1} + \frac{R(n_1-1)}{n_1+n_2}\right]$
$e^{-x} \quad 0 \leq x < \infty$ $0 \quad x < 0$	$1-(n_1-1)(n_2-1)\sum_{i=0}^{n_1-2}\sum_{j=0}^{n_2-2}\binom{n_1-2}{i}\binom{n_2-2}{j}\cdot\frac{(-1)^{i+j}}{[1+j+(1+i)R]\,(1+i)}$
$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \quad -\infty < x < \infty$	$n_1 = 2, \quad n_2 = 2 \quad \frac{2}{\pi}\tan^{-1}R$ $n_1 = 2, \quad n_2 = 3 \quad \frac{6}{\pi}\tan^{-1}\frac{R}{\sqrt{4+3R^2}}$ $n_1 = 3, \quad n_2 = 2 \quad \frac{6}{\pi}\left(\tan^{-1}\sqrt{3+4R^2} - \frac{\pi}{3}\right)$ $n_1 = 3, \quad n_2 = 3$ $\int_0^R dr\left[\frac{27r}{2\pi^2}\left\{\frac{2}{r^2}(u\tan^{-1}u - v\tan^{-1}v) + \frac{1}{6r^2(1+r^2)}(w\tan^{-1}w - u\tan^{-1}u) + \frac{u^2y}{r}\tan^{-1}2ry\right\}\right]$ where $u = [3(r^2+1)]^{-\frac{1}{2}} \qquad w = (7r^2+3)^{-\frac{1}{2}}$ $v = (4r^2+3)^{-\frac{1}{2}} \qquad y = (3r^2+4)^{-\frac{1}{2}}$

measure of the relative performance of these two tests seems desirable. An attempt to measure the performance of this test relative to the F test was made by comparing the tolerance intervals of the distribution of this ratio with those of the F test.

The length of the interval containing the central $1 - 2\alpha$ proportion of the distribution of F was compared with a similar length for the distribution of w_1/w_2 for $n_1 = n_2 = n$. The square of the ratio of these lengths will be called δ_α^2.

TABLE 2

$$Pr\left(\frac{w_1}{w_2} < R\right) = .005$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.0078	.0052	.0043	.0039	.0038	0037	.0036	.0035	.0034
3	.096	.071	.059	054	.051	.048	.045	.042	.041
4	.21	.16	.14	.13	.12	.12	.11	.11	.10
5	.30	.24	.22	20	.19	.18	.18	.17	.16
6	.38	.32	.28	.26	.25	.24	.23	.22	.22
7	.44	.38	.34	32	.30	.29	.28	.27	.26
8	.49	.43	.39	.36	.35	.33	.32	.31	.30
9	.54	.47	.43	.40	.38	37	.36	.35	.34
10	.57	.50	.46	.44	.42	.40	.39	.38	.37

$$Pr\left(\frac{w_1}{w_2} < R\right) = .01$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.0157	.0105	.0080	.0070	.0068	.0066	.0063	.0062	.0061
3	.136	.100	.084	.079	.073	.069	.065	.062	.060
4	.26	.20	.18	.17	.16	.15	.14	.14	.13
5	.38	.30	.26	.24	.23	.22	.21	.21	.20
6	.46	37	.33	.31	.29	.28	.27	.26	.26
7	.53	.43	.39	.36	.34	.33	.32	.31	.30
8	.59	.49	.44	.41	.39	.37	.36	.35	.34
9	.64	.53	48	.45	.43	.41	.40	.39	.38
10	.68	.57	.52	.49	.46	.45	.43	.42	.41

$$Pr\left(\frac{w_1}{w_2} < R\right) = .025$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.039	.026	.019	.018	.017	.016	.016	.015	.015
3	.217	.160	.137	.124	.115	.107	.102	.098	.095
4	.37	.28	.25	.23	.21	.20	.19	.18	.18
5	.50	.39	.34	.32	.30	.28	.27	.26	.25
6	.60	.47	.42	.38	.36	.34	.33	.32	.31
7	.68	.54	.48	.44	.42	.40	.38	.37	.36
8	.74	.59	.53	.49	.46	.44	.43	.42	.41
9	.79	.64	.57	.53	.50	.48	.47	.46	.44
10	.83	.68	.61	.57	.54	.52	.50	.49	.48

TABLE 2—*Continued*

$$Pr\left(\frac{w_1}{w_2} < R\right) = .05$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.079	.052	.039	.036	.034	.032	.031	.030	.028
3	.31	.23	.20	.18	16	.15	.14	.14	.13
4	.50	37	32	.29	.27	.26	.25	.24	.23
5	62	49	.42	.40	.36	.35	.33	.32	.31
6	.74	.57	.50	.46	.43	.41	.40	.38	.37
7	.80	.64	.57	.52	.49	47	.45	.44	.43
8	.86	.70	.62	57	.54	51	.50	.48	47
9	.91	75	.67	.61	.58	.55	53	52	.51
10	.95	.80	.70	.65	.61	.59	57	.55	.54

$$Pr\left(\frac{w_1}{w_2} < R\right) = .10$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.158	.105	.077	.074	.069	.066	.062	.059	.056
3	.46	.33	.28	.25	.23	.22	.21	.20	.19
4	.67	.49	.42	.38	.36	.34	.32	.31	.30
5	.84	.62	.53	48	.45	.43	.41	.39	.38
6	.97	.72	.62	.56	52	.50	.48	.46	.45
7	1 07	80	.69	.63	.59	.56	.54	.52	.50
8	1.15	87	.75	68	.64	.61	.58	.56	.54
9	1 21	92	.80	.73	.68	.65	.62	.60	.58
10	1.26	.98	.85	.77	.72	.68	.66	.64	.62

TABLE 3

n	*Relative precision of the range as an estimate of* σ	δ^2
2	1.00	1.00
3	.99	.99
4	.98	.97
5	.96	.95
6	.93	.92
7	.91	.90
8	.89	.89
9	.87	.88
10	.85	.86

For statistics having normal sampling distributions such a ratio would be independent of α and would be equivalent to the ratio of the variances of these sampling distributions. It was found that δ_α^2 is independent of α except for a maximum change of 1 in the second decimal for the values of α = .005, .01, .025, .05, .10. These values of δ^2 are presented in Table 3 along with the relative precision of the range as an estimate of σ as given by Mosteller [1].

It is interesting to note that δ^2 corresponds very closely to the relative precision of the range as an estimate of σ.

REFERENCE

[1] F. MOSTELLER, "On some useful 'inefficient' statistics," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 377–408.

A NOTE ON THE ESTIMATION OF A DISTRIBUTION FUNCTION BY CONFIDENCE LIMITS

BY FRANK J. MASSEY, JR.

University of Oregon

Let $F(x)$ be the continuous, cumulative distribution function of a random variable X, and let $x_1 < x_2 < x_3 < \cdots < x_n$ be the results of n independent observations on X arranged in order of size. We wish to estimate $F(x)$ by means of the band $S_n(x) \pm \lambda/\sqrt{n}$ where $S_n(x)$ is defined by

$$S_n(x) = \begin{cases} 0 & \text{if } x < x_1, \\ k/n & \text{if } x_k \le x < x_{k+1}, \\ 1 & \text{if } x \ge x_n. \end{cases}$$

Thus we wish to know the probability, say $P_n(\lambda)$, that the band is such that $S_n(x) - \dfrac{\lambda}{\sqrt{n}} < F(x) < S_n(x) + \dfrac{\lambda}{\sqrt{n}}$ for all x. This problem has been previously studied [1] [2] [3] [4] [5] and a limiting distribution has been obtained [1] [4] [5] and tabled [3] [4]. However apparently no error terms for the limiting distribution, or practical methods of obtaining $P_n(\lambda)$ have been given. Such a method is given here.

It has been shown [2] that $P_n(\lambda)$ is independent of $F(x)$ provided only that $F(x)$ is continuous, and thus it is sufficient to consider only the case

$$F(x) = \begin{cases} 0 \text{ if } x < 0, \\ x \text{ if } 0 \le x \le 1, \\ 1 \text{ if } x \ge 1. \end{cases}$$

We will find the probability that $S_n(x)$ falls wholly in the band $F(x) \pm k/n$ (here $\lambda = k/\sqrt{n}$) where k is an integer or a rational number, and intermediate values may be obtained by interpolation. To illustrate the method we shall assume that k is an integer.

Divide the interval (0, 1) into n parts by the points $1/n, 2/n, \cdots, (n-1)/n$. The step function $S_n(x)$ rises by jumps of exactly $1/n$. Thus, in order to be inside the band at $x = i/n$, $S_n(x)$ would have to pass through exactly one of the lattice points whose ordinates are $(i-k+1)/n, (i-k+2)/n, \cdots, (i+k-1)/n$.

Suppose that the step function stays inside the band by means of α_i of the observations falling in the interval $\left(\frac{i-1}{n}, \frac{i}{n}\right)$ $i = 1, 2, \cdots, n$. The a priori probability of this happening is given by the multinomial law as

$$P_r(\alpha_1 \cdots \alpha_n) = \frac{n!}{\alpha_1!\, \alpha_2! \cdots \alpha_n!} \left(\frac{1}{n}\right)^{\alpha_1} \left(\frac{1}{n}\right)^{\alpha_2} \cdots \left(\frac{1}{n}\right)^{\alpha_n}$$
$$= \frac{1}{\alpha_1! \cdots \alpha_n!} \frac{n!}{n^n}$$

since $\sum_1^n \alpha_i = n$.

Thus the probability of the step function staying in the band is given by

$$P_n(\lambda) = \sum \frac{n!}{n^n} \frac{1}{\alpha_1!\, \alpha_2! \cdots \alpha_n!} = \frac{n!}{n^n} \sum \frac{1}{\alpha_1! \cdots \alpha_n!}$$

where the summation is over all possible combinations of $\alpha_1, \cdots, \alpha_n$ such that $\max_x | S_n(x) - x | < \frac{\lambda}{\sqrt{n}}$ and $\sum_{i=1}^{n} \alpha_i = n$.

Let $U_i(m) = \sum_i \frac{1}{\alpha_1! \cdots \alpha_m!}$, $i = 1, 2, \cdots, 2k-1$ be the sum of all the terms indicated such that $S_n(x)$ arrives at the lattice point $\left(\frac{m}{n}, \frac{m-k+i}{n}\right)$ by a route that stays inside the band. Since the $S_n(x)$ is non-decreasing it can only pass through a point

$$\left(\frac{m+1}{n}, \frac{m-k+1+j}{n}\right), \quad m = 0, 1, \cdots, n-1; \qquad j = 1, 2, \cdots, 2k-1,$$

if it previously passed through one of the points

$$\left(\frac{m}{n}, \frac{m-k+1}{n}\right) \cdots, \left(\frac{m}{n}, \frac{m-k+2}{n}\right) \cdots, \left(\frac{m}{n}, \frac{m-k+j+1}{n}\right).$$

If it passed through $\left(\frac{m}{n}, \frac{m-k+h}{n}\right)$ the value of α_{m+1} would have to be $(j+1-h)$ and the product $U_h(m) \frac{1}{(j+1-h)!}$ would be part of $U_j(m+1)$. This is true for all $h = 1, 2, \cdots, j+1$ and all of these terms would give different paths for $S_n(x)$ so we have

$$U_j(m+1) = \sum_{h=1}^{j+1} \frac{1}{(j+1-h)!} U_h(m), \qquad j = 1, 2, \cdots, 2k-1,$$

where it is understood $U_h(m) = 0$ if $h \geq m + k$.

Thus we have a set of $2k - 1$ linear homogeneous difference equations. They may be reduced to a single difference equation by eliminating $2k - 2$ of the variables by substitution. This results in the following difference equation.

$$\sum_{h=1}^{2k-1} (-1)^h \frac{(2k-h)^h}{h!} U_k(2k - 1 - h + m) = 0.$$

TABLE 1

k	$n = 5$	10	20	25	30	35	40	45
1.0	.0384	.0004						
1 5	.3276	.0449						
2.0	6521	.2513	.0238					
2 5	.8880	5139						
3.0	.9699	.7331	.2955					
3.5	9947	.8522						
4.0	.99935	.9410	.6473					
5.0		.9922	.8624	.7637	.6629	.5674	.4808	.4042
6.0		.9994	.9569	.9057	.8420	.7725	.7016	.6322
7.0			.9892	.9683	.9359	.8945	.8471	7962
8.0			.9979	9911	.9774	.9566	.9295	8974
9.0			.9997	.9979	.9931	.9842	9708	.9529

k	$n = 50$	55	60	65	70	75	80
5 0	.3377	.2807	.2324	.1918	.1577	.1294	.1060
6.0	5662	.5046	.4478	.3954	.3492	.3072	.2696
7 0	.7439	.6916	.6403	.5908	.5435	.4987	.4566
8.0	.8616	8234	7837	.7434	.7031	.6633	6244
9 0	.9312	9063	.8789	.8496	.8189	.7874	.7554

Initial conditions on either the simultaneous equations or on the single equation are

$$U_i(0) = 0 \text{ for } i \neq k,$$

$$U_k(0) = 1 \text{ for } i = k.$$

After values of $U_k(n)$ have been found the value of $P_n\left(\frac{k}{\sqrt{n}}\right)$ can be found by multiplying $U_k(n)$ by $\frac{n!}{n^n}$.

The values of $U_k(n)$ can be obtained numerically either from the simultaneous

equations or from the single equation. Table 1 was computed partly by numerical solution of the simultaneous equations above and partly by setting up similar equations connecting $U_t(x+5)$ to $U_t(x)$, $t = 1, 2, \cdots, i+5$. Either method could be set up on punch cards if an extensive table was desired. Notice that to get $U_k(n)$ all $U_k(t)$, $t = 1, 2, \cdots, n-1$ are also found. Table 1 gives some computed values of $P_n(k)$. Table 2 gives results interpolated from Table 1, showing the approach of $P_n(\lambda)$ to its limiting distribution.

If the width of the band is $2\left(\frac{k}{l}\right)$ when k and l are integers a similar procedure to that above can be used. However instead of dividing the interval (0, 1) into n parts it is necessary to divide it into $l \cdot n$ parts.

TABLE 2

n	$\lambda = .9$	1.0	1.10	1.20	1.30	1.40
10	.66	.78	.85	.91	.95	.97
20	.65	.77	.85	.91	.94	.97
30	.65	.76	.85	.90	.94	.96
40	.64	.76	84	.90	.94	.96
50	.64	.75	.84			
60	.63	.75	.84			
70	.63	.75	.83			
80	.63	.74				
∞	.607	.730	.822	.888	.932	.960

It has been suggested (2) that instead of a band bounded by $y = x \pm c$ it might be convenient to use a band bounded by the lines $y = px + q$ and $y = p'x + q'$. If $p = p'$ and if p, q, q' are rational the probability of $S_n(x)$ staying inside the band can be evaluated by the method presented above. If $p \neq p'$ and if p, p', q, q' are all rational a similar procedure could be used but it would be very tedious.

[1] N. Smirnov, "Sur les écarts de la courbe de distribution empirique," *Recueil Math. de Moscou*, Vol. 6 (1939), pp. 3–26.
[2] A. Wald and J. Wolfowitz, "Confidence limits for continuous distribution functions," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 105–118.
[3] N. Smirnov, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bulletin Mathematique de l'Universite de Moscou*, Vol. 2 (1939), fasc. 2.
[4] A. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *Ist. Ital. att.* Giorn. Vol. 4 (1933), p. 1–11.
[5] W. Feller, "On the Kolmogorov-Smirnov limit theorems for empirical distribution" *Annals of Math. Stat.*, Vol. 19 (1948), pp. 177–190.

SIGNIFICANCE LEVELS FOR A k-SAMPLE SLIPPAGE TEST[1]

FREDERICK MOSTELLER AND JOHN W. TUKEY

Harvard University and Princeton University

1. Summary. Mosteller has recently [1, 1948] proposed a k-sample slippage test and has given percentage points for selected n, k and r for the case of k equal samples of size n. When the samples are of unequal size, exact significance levels can be calculated very quickly from

$$P_r = \frac{\sum n_i^{(r)}}{N^{(r)}} \text{ where } x^{(r)} = x(x-1)\cdots(x-r+1),$$

by the method explained in section 3 below.

The significance values for k equal samples of $n \geq 10$ are very well approximated by

$$P_r = \frac{1}{k^{r-1}} e^{-r(r-1)(k-1)/2N}$$

where $N = kn$.

A convenient rough approximation for unequal samples may be given in terms of k^*, an "effective" number of samples, which is given by

$$k^* = \frac{(\sum_i n_i)^2}{\sum n_i^2},$$

the one-sided significance level will then be approximately given by

$$P_r = (k^*)^{-(r-1)}.$$

This approximation can be easily applied with the aid of Table 1. Thus, for example, with four samples of sizes 7, 5, 5, 2, we have

$$k^* = \frac{(7+5+5+2)^2}{49+25+25+4} = \frac{361}{103} = 3.50,$$

whence from the table $r = 3$ lies at a one-sided level approximately between 5% and 10%, $r = 4$ approximately between 1% and 2.5%, $r = 5$ between 0.5% and 1%, $r = 6$ near 0.2%, and so on. Direct calculation yields 5.7%, 1.2%, 0.2% and 0.03%. The approximation is, in this example, quite satisfactory for moderate significance levels and conservative for more extreme significance levels.

2. Derivation. The statistic considered by Mosteller is the number of cases in one sample greater than all cases in all the $k-1$ other samples. We derive its distribution briefly.

Since the statistic depends only on the order of the $n_1 + n_2 + \cdots + n_k = N$

[1] Prepared in connection with research sponsored by the Office of Naval Research.

values, we can consider the actual values taken on to be fixed, and consider their allotment to the various samples. Assuming all of them to come from a single continuous distribution, we may consider these fixed values to be all distinct, and any way of allotting them to labelled places in the various samples as equally likely.

Consider the r largest values. They can all be allotted to places in the i-th sample in $n_i^{(r)} = n_i(n_i - 1) \cdots (n_i - r + 1)$ ways, and to arbitrary places in $N^{(r)}$ ways. Thus they will be allotted to some single sample in the fraction

$$P_r = \frac{\sum_i n_i^{(r)}}{N^{(r)}}$$

of all cases. This is clearly the probability that Mosteller's statistic is r or more.

TABLE 1

Approximate critical values of k^* for various levels of significance

One-sided level	10%	5%	2.5%	1%	0.5%	0.2%	0.1%
Two sided level	20%	10%	5%	2%	1%	0.4%	0.2%
$r = 2$	10.0	20.0	40.0	100.0	200.0	500.0	1000.0
$r = 3$	3.2	4.5	6.3	10.0	14.1	22.4	31.6
$r = 4$		2.7	3.4	4.6	5.8	7.9	23.0
$r = 5$			2.5	3.2	3.8	4.7	5.6
$r = 6$				2.5	2.9	3.5	4.0
$r = 7$						2.8	3.2
$r = 8$							2.6

3. Unequal samples—an exact computation. Our practical problem is to compute P_r for small values of r and a fixed set of n_i. If we recognize the numerators as the unnormalized factorial moments of the distribution of sample sizes, we see that the computation goes smoothly according to the scheme shown in Table 2 (where the columns of multipliers $n - 1, n - 2, n - 3$, etc. may be partially covered for convenience during the computation.): For example: $132 = 11(12)$, $1320 = 10(132)$, $\cdots 42 = 6(7)$. The numbers in the last line of Table 2 give successively the percentages $100\, P_1$, $100\, P_2$, $\cdots$. Of course $P_1 = 1$ because some sample must have the largest value. It is clear that exact computation for any reasonable set of n_i is quite easy.

4. Equal samples—an approximation. In the case of k equal samples, we have

$$P_r = \frac{kn^{(r)}}{N^{(r)}}.$$

Let us try to approximate to $n^{(r)}$ by expansion in powers. We have

$n^{(r)} = n(n-1)\cdots(n-r+1) = n^r(1-1/n)(1-2/n)\cdots(1-(r-1)/n),$
so that

$$\begin{aligned}\log n^{(r)} &= r\log n + \sum_{x=1}^{r-1}\log(1-x/n)\\ &= r\log n - \sum_{x=1}^{r-1}(x/n + x^2/2n^2 + x^3/3n^3\cdots)\\ &= r\log n - r(r-1)/2n - r(r-1)(2r-1)/12n^2 + 0(n^{-3}),\end{aligned}$$

TABLE 2

Sample Computation

for $\{n_i\}$ = (12, 11, 11, 11, 10, 10, 10, 10, 9, 9, 7, 4)

$n-2$	$n-1$	n	f	nf	$n^{(2)}f$	$n^{(3)}f$
10	11	12	1	12	132	1320
9	10	11	3	33	330	2970
8	9	10	4	40	360	2880
7	8	9	2	18	144	1008
5	6	7	1	7	42	210
2	3	4	1	4	12	24
$N-2$	$N-1$	N	Sums	114	1020	8412
112	113	114	$N^{(r)}$	114	12882	1442784
			P_r	100%	7.9%	0.58%

and hence

$$\begin{aligned}\log P_r &= \log k + \log n^{(r)} - \log N^{(r)} = \log k + \log n^{(r)} - \log (nk)^{(r)}\\ &= \log k + r\log n - r(r-1)/2n - r(r-1)(2r-1)/12n^2\\ &\qquad - r\log nk + r(r-1)/2nk + r(r-1)(2r-1)/12n^2k^2 + 0(n^{-3})\\ &= -(r-1)\log k - \frac{r(r-1)}{2n}\left(1 - \frac{1}{k} + \frac{2r-1}{6n} - \frac{2r-1}{6nk^2} + 0(n^{-2})\right).\end{aligned}$$

We get the following three approximations:

(1) $$P_r \doteq \frac{1}{k^{r-1}};$$

and noting that $\dfrac{1-1/k}{n} = \dfrac{k-1}{kn} = \dfrac{k-1}{N}$,

(2) $$P_r \doteq k^{-(r-1)}\, e^{(-r(r-1)(k-1))/2N} = \frac{1}{[ke^{(r(k-1))/2N}]^{r-1}};$$

and finally

$$P_r \doteq k^{-(r-1)} e^{(-r(r-1)(k-1)/2N)(1+(2r-1)/6n)}. \tag{3}$$

5. Comparison of results. The results obtained with various equal sample approximations will be compared with the exact values for several cases. The effective number of samples, k^*, used with (1), (2), and (3), is computed from

$$k^* = \frac{(\sum n_i)^2}{\sum n_i^2},$$

a formula which is often an easy and effective way to allow for different sizes of samples.

TABLE 3

Comparison of Approximations

Sizes of Samples	N	k	r	exact	P_r in (1)	(2)	(3)	(4)
10, 10, 10, 10	40	4.00	2	23.08	25.00	23.19	23.13	$\leq$25.00
			3	4.85	6.25	4.99	4.80	$\leq$6.25
7, 5, 5, 2	19	3.50+	2	24.56	28.53	25.01	24.82	$\leq$28.53
			3	5.67	8.14	5.48	5.18	$\leq$8.76
12, 11, 11, 11 10, 10, 10, 10	114	11.46	2	7.92	8.73	7.96	7.96	$\leq$8.73
9, 9, 7, 4			3	0.58	0.76	0.58	0.56	$\leq$0.78

A fourth approximation, which always gives a conservative estimate of the significance of the result is obtained by replacing $n^{(r)}$ by n^r throughout, this gives

$$P_r = \frac{\sum n_i^r}{N^r}, \tag{4}$$

which is equivalent to approximation (1) when the samples are of equal size, or when $r = 2$.

The results are shown in Table 3.

Thus it seems clear that either (1) or (4) are good enough for rough work. The choice will depend on which formula one prefers to remember. The amount of work is about the same for either method. When something better is required the exact method of section 3 seems appropriate. Indeed some may prefer it to any approximation.

REFERENCE

[1] FREDERICK MOSTELLER, "A k-sample slippage test for an extreme population," *Annals of Math. Stat.*, Vol. 19 (1948), p. 58–65.

ADJUSTMENT OF AN INVERSE MATRIX CORRESPONDING TO A CHANGE IN ONE ELEMENT OF A GIVEN MATRIX

By Jack Sherman and Winifred J. Morrison

The Texas Company Research Laboratories, Beacon, New York

1. Introduction. Many methods have been published in recent years for carrying out the numerical computation of the inverse of a matrix [1], [2]. In all these methods, the amount of computation increases rapidly with increase in order of the matrix

The utility of a computational method for obtaining the inverse of a matrix would be increased considerably if the inverse could be transformed in a simple manner, corresponding to some specified change in the original matrix, thus eliminating the necessity of computing the new inverse from the beginning. The problem that is considered in the present paper is one of changing one element in the original matrix, and of computing the resulting changes in the elements of the new inverse directly from those of the old inverse.

2. Computational method. Let

a_{ij}, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, n$ denote the elements of an nth order square matrix $\boldsymbol{a}$;

b_{ij}, denote the elements of $\boldsymbol{b}$, the inverse of $\boldsymbol{a}$;

A_{ij}, denote the elements of $\boldsymbol{A}$ which differs from $\boldsymbol{a}$ only in one element, say A_{RS};

B_{ij}, denote the elements of $\boldsymbol{B}$, the inverse A.

Let

$$A_{RS} = a_{RS} + \Delta a_{RS}.$$

The set of equations by means of which $\boldsymbol{B}$ may be computed from Δa_{RS} and $\boldsymbol{b}$ is

$$B_{rj} = b_{rj} - \frac{b_{rR}\, b_{Sj}\, \Delta\, a_{RS}}{1 + b_{SR}\, \Delta\, a_{RS}}, \qquad r = 1, 2, \cdots, n, \quad j = 1, 2, \cdots, n, \tag{1}$$

provided that $1 + b_{SR}\Delta a_{RS} \neq 0$.

The validity of equation (1) may be demonstrated by multiplying through by A_{ir}, $(r = 1, 2, \cdots, n)$ and adding the results:

$$\sum_{r=1}^{n} A_{ir} B_{rj} = \sum_{r=1}^{n} A_{ir}\, b_{rj} - \frac{b_{Sj}\, \Delta\, a_{RS}}{1 + b_{SR}\, \Delta\, a_{RS}} \sum_{r=1}^{n} A_{ir}\, b_{rR}, \tag{2}$$

$$(i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, n).$$

Consider separately the equations for which $i \neq R$, and for which $i = R$.

Case I. $i \neq R$. By hypothesis, $A_{ir} = a_{ir}$ for $i \neq R$. Hence equations (2) become

$$\sum_{r=1}^{n} A_{ir} B_{rj} = \sum_{r=1}^{n} a_{ir}\, b_{rj} - \frac{b_{Sj}\, \Delta\, a_{RS}}{1 + b_{SR}\, \Delta\, a_{RS}} \sum_{r=1}^{n} a_{ir}\, b_{rR}, \tag{3}$$

$$(i = 1, 2, \cdots, R-1, R+1, \cdots, n;\ j = 1, 2, \cdots, n).$$

The last sum vanishes because a and b are inverse matrices, and hence

$$\sum_{r=1}^{n} A_{ir} B_{rj} = \sum_{r=1}^{n} a_{ir} b_{rj} \tag{4}$$
$$(i = 1, 2, \cdots, R-1, R+1, \cdots, n; j = 1, 2, \cdots, n).$$

Case II. $i = R$. Equation (2) becomes

$$\sum_{r=1}^{n} A_{Rr} B_{rj} = \sum_{r=1}^{n} A_{Rr} b_{rj} - \frac{b_{Sj} \Delta a_{RS}}{1 + b_{SR} \Delta a_{RS}} \sum_{r=1}^{n} A_{Rr} b_{rR} \quad (j = 1, 2, \cdots, n) \tag{5}$$

In each of the summations, there will be a term for which $r = S$, in which case $A_{RS} = a_{RS} + \Delta a_{RS}$. In all other cases, $A_{Rr} = a_{Rr}$. Hence (5) can be written as

$$\begin{aligned} \sum_{r=1}^{n} A_{Rr} B_{rj} &= \sum_{r=1}^{n} a_{Rr} b_{rj} + \Delta a_{RS} b_{Sj} \\ &\quad - \left(\frac{b_{Sj} \Delta a_{RS}}{1 + b_{SR} \Delta a_{RS}}\right)\left(\sum_{r=1}^{n} a_{Rr} b_{rR} + \Delta a_{RS} b_{SR}\right) \qquad (j = 1, 2, \cdots, n). \end{aligned} \tag{6}$$

Since a and b are inverse matrices, the second summation on the right-hand side of (6) is equal to unity, and hence (6) becomes

$$\sum_{r=1}^{n} A_{Rr} B_{rj} = \sum_{r=1}^{n} a_{Rr} b_{rj} \qquad (j = 1, 2, \cdots, n). \tag{7}$$

The sets of equations (4) and (7) can be written as one set of equations:

$$\sum_{r=1}^{n} A_{ir} B_{rj} = \sum_{r=1}^{n} a_{ir} b_{rj} \quad (i = 1, 2, \cdots, n; j = 1, 2, \cdots, n), \tag{8}$$

and hence B is the inverse of A.

3. Illustrative numerical example. In actual applications, equations (1) are conveniently subdivided into three groups, namely, those for which $r = S$, those for which $j = R$, and all others. In the first two cases, these reduce to

$$B_{Sj} = \frac{b_{Sj}}{1 + b_{SR} \Delta a_{RS}}, \qquad (j = 1, 2, \cdots, n), \tag{9}$$

$$B_{rR} = \frac{b_{rR}}{1 + b_{SR} \Delta a_{RS}}, \qquad (r = 1, 2, \cdots, n). \tag{10}$$

By utilizing (10), (1) becomes

$$\begin{aligned} &B_{rj} = b_{rj} - B_{rR} b_{Sj} \Delta a_{RS}, \\ &(r = 1, 2, \cdots, S-1, S+1, \cdots, n; \\ &j = 1, 2, \cdots, R-1, R+1, \cdots, n). \end{aligned} \tag{11}$$

Equations (10) and (11) show that the elements of B contained in the Sth row and Rth column are directly proportional to the corresponding elements of b.

Consider

$$a = \begin{pmatrix} 2.384 & 1.238 & 0\ 861 & 2.413 \\ 0.648 & 1.113 & 0.761 & 0\ 137 \\ 1.119 & 0.643 & 3.172 & 1.139 \\ 0.745 & 2.137 & 1.268 & 0.542 \end{pmatrix}.$$

The inverse of b turns out to be

$$b = \begin{pmatrix} 0.2220 & 2.5275 & -0.1012 & -1.4145 \\ -0.04806 & -0.2918 & -0.1999 & 0.7079 \\ -0.1692 & 0.01195 & 0.3656 & -0.01824 \\ 0.2801 & -2.3517 & 0.07209 & 1.0409 \end{pmatrix}.$$

Assume that a_{24} is increased by 0.4, so that

$$A = \begin{pmatrix} 2.384 & 1.238 & 0.861 & 2.413 \\ 0.648 & 1.113 & 0.761 & 0.537 \\ 1.119 & 0.643 & 3.172 & 1.139 \\ 0\ 745 & 2.137 & 1.268 & 0.542 \end{pmatrix}.$$

Then (9), (10), and (11) become

$$B_{4j} = \frac{b_{4j}}{1 - 2.3517 \times 0.4} = 16.857\, b_{4j} \qquad (j = 1, 2, \cdots, n),$$

$$B_{r2} = 16.857\, b_{r2}\ (r = 1, 2, \cdots, n),$$

$$B_{rj} = b_{rj} - 0.4\, B_{r2} b_{4j}\ (r = 1, 2, \cdots, S-1, S+1, \cdots, n;$$
$$j = 1, 2, \cdots, R-1, R+1, \cdots, n).$$

Utilization of these equations gives

$$B = \begin{pmatrix} -4.5518 & 42.608 & -1.3298 & -19.155 \\ 0.5031 & -4.9191 & -0.05805 & 2.7560 \\ -0.1919 & 0.2014 & 0.3598 & -0.1021 \\ 4.7218 & -39.644 & 1.2153 & 17.547 \end{pmatrix}.$$

4. Concluding remarks. It is seen from equation (1) that if $\Delta a_{RS} = -1/b_{SR}$, that is, if a_{RS} is increased by the negative of the reciprocal of the corresponding element in the transposed reciprocal matrix, then the denominator in the second term on the right-hand side of equation (1) becomes equal to zero, and B cannot be found by the present method. It is left to the reader to verify that under these conditions A is in fact singular.

In the illustrative numerical example, the denominator is only $1 - 2.3517 \times 0.4 = 0.05932$, which accounts for the large magnitude of some of the elements of B. If Δa_{24} were taken to be $1/2.3517 = 0.4252$ instead of 0.4, A would have become singular.

If two or more elements in the matrix a are to be changed, the new inverse can be found by successive applications of the method.

REFERENCES

[1] H. Hotelling, "Some new methods in matrix calculation," *Annals of Math Stat.*, Vol. 14 (1943), pp. 1–35.

[2] P. S. Dwyer, "The solution of simultaneous equations," *Psychometrika*, Vol. 6 (1941), p. 101.

A CLASS OF RANDOM VARIABLES WITH DISCRETE DISTRIBUTIONS

Albert Noack

Cologne, Germany

1. General results. A large class of random variables with discrete probability distributions can be derived from certain power series. Let

$$f(z) = \sum_{x=0}^{\infty} a_x z^x, \quad a_x \text{ real}, \quad |z| < r.$$

We may have either non-negative coefficients a_x or we may have $(-1)^x a_x \geq 0$. In the first case take $0 < z < r$; and in the second case take $-r < z < 0$. Define a random variable with the distribution

$$P\{\xi = x\} = \frac{a_x z^x}{f(z)}; \qquad x = 0, 1, 2, \cdots . \tag{1}$$

The above conditions insure $P\{\xi = x\} \geq 0$ for all x; besides

$$\sum_x P\{\xi = x\} = \frac{1}{f(z)} \sum_x a_x z^x = 1.$$

The distribution of ξ may be called the power series distribution (p.s.d.).
The mean of such a distribution is

$$E(\xi) = \sum_x xP\{\xi = x\} = \frac{1}{f(z)} \sum_x x a_x z^x.$$

Hence it follows that

$$E(\xi) = z\frac{f'(z)}{f(z)} = z\frac{d}{dz} \log f(z). \tag{2}$$

We have for the moments about the origin

$$\mu_r' = \sum_x x^r P\{\xi = x\} = \frac{1}{f(z)} \sum_x x^r a_x z^x,$$

and hence

$$z\frac{d\mu_r'}{dz} = \frac{1}{f(z)} \sum_x x^{r+1} a_x z^x - z\frac{f'(z)}{f(z)} \frac{1}{f(z)} \sum_x x^r a_x z^x.$$

Thus we have the recurrence relation

$$\mu'_{r+1} = z\frac{d\mu'_r}{dz} + \mu'_1\mu'_r . \tag{3}$$

The central moments are

$$\mu_r = \sum_x (x - \mu'_1)^r P\{\xi = x\} = \frac{1}{f(z)} \sum_x (x - \mu'_1)^r a_x z^x,$$

and hence

$$z\frac{d\mu_r}{dz} = \frac{1}{f(z)} \sum_x x(x - \mu'_1)^r a_x z^x - zr\frac{d\mu'_1}{dz}\frac{1}{f(z)} \sum_x (x - \mu'_1)^{r-1} a_x z^x$$
$$- z\frac{f'(z)}{f(z)}\cdot\frac{1}{f(z)} \sum_x (x - \mu'_1)^r a_x z^x.$$

The sum of the first and third term will be found to be μ_{r+1}, hence

$$z\frac{d\mu_r}{dz} = \mu_{r+1} - rz\frac{d\mu'_1}{dz}\mu_{r-1},$$

whence we have for the central moments of a p.s.d. the recurrence relation

$$\mu_{r+1} = z\left[\frac{d\mu_r}{dz} + r\frac{d\mu'_1}{dz}\mu_{r-1}\right]. \tag{4}$$

Putting $r = 1$, $\mu_0 = 1$, $\mu_r = 0$, we get the variance of ξ

$$\mu_2 = \sigma^2(\xi) = z\frac{d\mu'_1}{dz} = z^2\frac{d^2}{dz^2}\log f(z) + \mu'_1 = z^2\frac{f''(z)}{f(z)} - z^2\left[\frac{f'(z)}{f(z)}\right]^2 + z\frac{f'(z)}{f(z)}. \tag{5}$$

By (5), (4) assumes the form

$$\mu_{r+1} = z\frac{d\mu_r}{dz} + r\mu_2\mu_{r-1} . \tag{4'}$$

The characteristic function of ξ is

$$\varphi(t) = \sum_x e^{itx} P\{\xi = x\} = \frac{1}{f(z)} \sum_x a_x e^{itx} z^x,$$

or

$$\varphi(t) = \frac{f(e^{it}z)}{f(z)}. \tag{6}$$

To get a relation connecting the cumulants κ_n and the moments μ'_r about the origin, we differentiate both sides of the identity

$$\sum_{r=1}^{\infty} \frac{\kappa_r}{r!}(it)^r = \log \sum_{\rho=0}^{\infty} \frac{\mu'_\rho}{\rho!}(it)^\rho$$

with respect to (it), identifying coefficients in $(it)^{r-1}$ we get[1]

$$\mu'_r = \sum_{j=1}^{r} \binom{r-1}{j-1} \mu'_{r-j} \kappa_j . \tag{7}$$

Differentiation of (7) with respect to z gives

$$\frac{d\mu'_r}{dz} = \sum_{j=1}^{r} \binom{r-1}{j-1} \left[\frac{d\mu'_{r-j}}{dz} \kappa_j + \mu'_{r-j} \frac{d\kappa_j}{dz} \right]. \tag{7'}$$

Substitution of (7) and (7′) in (3) gives

$$\sum_{j=1}^{r+1} \binom{r}{j-1} \mu'_{r+1-j} \kappa_j = \sum_{j=1}^{r} \binom{r-1}{j-1} \left\{ \left[z \frac{d\mu'_{r-j}}{dz} + \mu'_1 \mu'_{r-j} \right] \kappa_j + z\mu'_{r-j} \frac{d\kappa_j}{dz} \right\},$$

or by (3) after a little re-arrangement

$$\kappa_{r+1} = z \sum_{j=1}^{r} \binom{r-1}{j-1} \mu'_{r-j} \frac{d\kappa_j}{dz} - \sum_{j=2}^{r} \binom{r-1}{j-2} \mu'_{r+1-j} \kappa_j . \tag{8}$$

2. Special cases.

(a) Choosing $f(z) = e^z$, ξ has Poisson-distribution

$$P\{\xi = x\} = \frac{z^x e^{-z}}{x!}. \tag{1a}$$

(2) and (5) are the well known relations $E(\xi) = \sigma^2(\xi) = z$; the recurrence formula (4) assumes the form[2]

$$\mu_{r+1} = z \left[\frac{d\mu_r}{dz} + r\mu_{r-1} \right]. \tag{4a}$$

(b) Taking $f(z) = (1 - z)^{-k}$, $k > 0$, $0 < z < 1$ we get the so-called negative binomial distribution

$$P\{\xi = x\} = \frac{\Gamma(k + x)}{x!\,\Gamma(k)} z^x (1 - z)^k, \qquad x = 0, 1, 2, \cdots . \tag{1b}$$

The mean is

$$E(\xi) = \frac{kz}{1 - z}, \tag{2b}$$

while the recurrence formula for the central moments is

$$\mu_{r+1} = z \left[\frac{d\mu_r}{dz} + \frac{rk}{(1 - z)^2} \mu_{r-1} \right], \tag{4b}$$

hence the first three moments of this distribution are

$$\sigma^2(\xi) = \mu_2 = \frac{kz}{(1 - z)^2},$$

[1] Cf. M. G. Kendall, *The Advanced Theory of Statistics*, Vol. I, p 87

[2] Cf. Craig, *Am. Math. Soc. Bull.*, Vol 40 (1934), p. 262.

$$\mu_3 = \frac{kz(1+z)}{(1-z)^3}, \tag{5b}$$

$$\mu_4 = \frac{kz(1+4z+z^2+3kz)}{(1-z)^4}.$$

The characteristic function of the distribution is

$$\varphi(t) = \left(\frac{1-e^{it}z}{1-z}\right)^{-k}. \tag{6b}$$

Writing $z = \eta/(1+\eta)$, $k = h/\eta$, $\eta > 0$, $h > 0$ we get the so-called Polya-Eggenberger distribution for rare contagious events[3].

$$w\{\xi = x\} = \frac{\Gamma\left(\frac{h}{\eta}+x\right)}{x!\,\Gamma(h\eta^{-1})}\left(\frac{\eta}{1+\eta}\right)^x(1+\eta)^{-h/\eta}, \qquad x = 0, 1, 2, \cdots. \tag{1b$_1$}$$

The first four moments of this distribution are

$$\mu_1' = h \tag{2b$_1$}$$

$$\mu_2 = h(1+\eta) \tag{5b$_1$}$$

$$\mu_3 = h(1+\eta)(1+2\eta)$$

$$\mu_4 = h(1+\eta)[1+3(1+\eta)(h+2\eta)].$$

To obtain a recurrence relation for the moments consider

$$\frac{d\mu_r}{dz} = \frac{\partial\mu_r}{\partial\eta}\frac{d\eta}{dz} + \frac{\partial\mu_r}{\partial h}\frac{dh}{\partial z} = (1+\eta)^2\left[\frac{\partial\mu_r}{\partial\eta} + \frac{h}{\eta}\frac{\partial\mu_r}{\partial h}\right];$$

hence we find for this distribution by (4) and (4b)

$$\mu_{r+1} = (1+\eta)\left[\eta\frac{\partial\mu_r}{\partial\eta} + h\frac{\partial\mu_r}{\partial h} + rh\mu_{r-1}\right]. \tag{4b$_1$}$$

It follows from (4b$_1$), that μ_r is a polynomial in η and h. The characteristic function of this distribution is

$$\varphi(t) = [1+\eta(1-e^{it})]^{-h/\eta}. \tag{6b$_1$}$$

(c) The coefficients of the series $-\log(1-z) = \sum_{x=1}^{\infty} z^x/x$ are positive; the associated distribution derived is

$$P\{\xi = x\} = -\frac{z^x}{x\log(1-z)}, \qquad 0 < z < 1; \quad x = 1, 2, \cdots, \tag{1c}$$

and has the mean

$$E(\xi) = -\frac{z}{(1-z)\log(1-z)}. \tag{2c}$$

[3] Cf *Zeits f. angew Math und Mech.*, Vol. 3 (1923), p. 279–289.

Recurrence formula (4) has for this distribution the form

$$\text{(4c)} \qquad \mu_{r+1} = z\left[\frac{d\mu_r}{dz} - r\,\frac{z + \log(1-z)}{(1-z)^2[\log(1-z)]^2}\,\mu_{r-1}\right],$$

while the variance and the characteristic function of this distribution are

$$\text{(5c)} \qquad \mu_2 = \sigma^2(\xi) = -\frac{z^2 + z\log(1-z)}{(1-z)^2[\log(1-z)]^2},$$

$$\text{(6c)} \qquad \varphi(t) = \frac{\log(1 - e^{it}z)}{\log(1-z)}.$$

(d) The coefficients of the series $\log(1+z)/(1-z) = 2\sum_{x=1}^{\infty}(z^{2x+1})/(2x+1)$ are positive, so we can derive a random variable ξ with the distribution

$$\text{(1d)} \quad P\{\xi = 2x + 1\} = \frac{2z^{2x+1}}{(2x+1)\log\dfrac{1+z}{1-z}}, \quad 0 < z < 1, x = 1, 2, 3, \cdots.$$

ξ has the mean

$$\text{(2d)} \qquad E(\xi) = \frac{2z}{(1-z^2)\log\dfrac{1+z}{1-z}},$$

the recurrence formula (4) assumes the form

$$\text{(4d)} \qquad \mu_{r+1} = z\left(\frac{d\mu_r}{dz} + 2r\cdot\frac{(1+z^2)\log\dfrac{1+z}{1-z} - 2z}{(1-z^2)^2\left[\log\dfrac{1+z}{1-z}\right]^2}\,\mu_{r-1}\right),$$

while the variance and the characteristic function of ξ are

$$\text{(5d)} \qquad \sigma^2(\xi) = 2z\,\frac{(1+z^2)\log\dfrac{1+z}{1-z} - 2z}{(1-z^2)^2\left[\log\dfrac{1+z}{1-z}\right]^2},$$

$$\text{(6d)} \qquad \varphi(t) = \frac{\log(1 + e^{it}z) - \log(1 - e^{it}z)}{\log(1+z) - \log(1-z)}.$$

(e) Likewise the coefficients of the series

$$\sin^{-1} z = z + \sum_{x=1}^{\infty}\frac{1\cdot 3\cdot 5\cdots(2x-1)}{2\cdot 4\cdot 6\cdots(2x)}\,\frac{z^{2x+1}}{2x+1}$$

are positive, the derived variable ξ with the distribution

$$P\{\xi = 1\} = (\sin^{-1} z)^{-1},$$

(1e) $$P\{\xi = 2x + 1\} = \frac{1\cdot 3\cdots(2x-1)}{2\cdot 4\cdot 6\cdots(2x)}\cdot\frac{z^{2x+1}}{2x+1}(\sin^{-1} z)^{-1},$$
$$0 < z < 1, x = 1, 2, 3, \cdots,$$

has the mean

(2e) $$E(\xi) = \frac{z}{\sqrt{1-z^2}\sin^{-1} z}.$$

The recurrence formula for the moments

(4e) $$\mu_{r+1} = z\left[\frac{d\mu_r}{dz} + r\,\frac{\sin^{-1} z - z\sqrt{1-z^2}}{\sqrt{1-z^2}^2(\sin^{-1} z)^2}\,\mu_{r-1}\right]$$

gives the variance

(5e) $$\sigma^2(\xi) = z\,\frac{\sin^{-1} z - z\sqrt{1-z^2}}{\sqrt{1-z^2}^3(\sin^{-1} z)^2}.$$

The characteristic function assumes the form

(6e) $$\varphi(t) = \frac{\sin^{-1} e^{it} z}{\sin^{-1} z}.$$

(f) It is well known, that series (b), (c), (d), and (e) are special cases of the hypergeometric function $F(a, b, c; z)$. This function gives a p.s.d., if $abc > 0$. If $a > 0, b > 0, c > 0$ or if $a < 0, b < 0, c > 0$, a, b integers, there exist no further restrictions on these parameters. Suppose $a < 0, b < 0, c > 0$, a integer, b not, we must have $[b] \leq a$[4]; if neither a nor b are integers, we must have $[a] = [b]$. Suppose $a < 0, b > 0, c < 0$. If c is an integer, a must be an integer $> c$. If a is an integer, but c not, we must have $[c] \leq a$. Finally if neither a nor c are integers, we must have $[a] = [c]$. Corresponding conditions are valid, if $a > 0, b < 0, c < 0$. Regarding

$$\frac{d}{dz}F(a, b; c; z) = \frac{ab}{c}F(a+1, b+1; c+1; z),$$

the mean of a random variable ξ with hypergeometric distribution is

(2f) $$E(\xi) = z\,\frac{ab}{c}\,\frac{F(a+1, b+1; c+1; z)}{F(a, b; c; z)}.$$

Considering the differential equation

$$z(1-z)f''(z) + [c - (a+b+1)z]f'(z) - abf(z) = 0,$$

(5) gives the variance of ξ

(5f) $$\sigma^2(\xi) = \frac{ab}{c}\cdot\frac{z}{1-z}\left\{c + [1 - c + (a+b)z]\,\frac{F(a+1, b+1; c+1; z)}{F(a, b; c; z)} - z(1-z)\,\frac{ab}{c}\left[\frac{F(a+1, b+1; c+1; z)}{F(a, b; c; z)}\right]^2\right\}.$$

The higher moments of this distribution can now derived from (4′).

[4] $[b]$ means as usual the greatest integer $\leq b$.

THE GEOMETRIC RANGE FOR DISTRIBUTIONS OF CAUCHY'S TYPE

BY E. J. GUMBEL AND R. D. KEENEY

New York City and Metropolitan Life Insurance Company

1. Introduction. We consider large samples drawn from a symmetrical unlimited population whose distribution is of the Cauchy type, defined by the properties

$$\lim_{x\to\infty} x^k[1 - F(x)] = A, \qquad \lim_{x\to-\infty} (-x)^k F(x) = A, \tag{1}$$

where k and A are positive and $F(x)$ stands for the probability function. This type of distribution has no moments of an order equal to or greater than k. We construct the distribution of a certain function of the extreme values, and require only the knowledge of the type of the initial distribution, not of the distribution itself.

From each sample we pick out the largest and smallest observations, x_n and x_1. If the median of the initial distribution is zero, and the sample size is large enough, the probability of any extreme x_n or $-x_1$ being negative can be neglected. If we draw N such samples, each of large size n, we obtain N pairs of extremes, $x_{n,\nu}$ and $x_{1,\nu}$ $(\nu = 1, 2, 3, \cdots, N)$. For each sample we can then compute the geometric mean, ρ, of these extremes:

$$\rho = \sqrt{x_n(-x_1)}, \tag{2}$$

which we henceforth call the *geometric* range.

The distribution of these geometric ranges can be obtained directly from the joint asymptotic distribution of the extremes. However, it is easier to obtain this distribution indirectly from the distribution of the reciprocal of the geometric range. This distribution of the reciprocal is of interest in itself: since it possesses all moments we can use it to estimate the parameters by the method of moments, whereas this problem seems to be very intricate if we start from the distribution of the geometric range itself.

2. The distribution of the reciprocal of the geometric range. The distribution of the reciprocal of the geometric range follows from a theorem of Elfving [1] which may be stated thus:

"Let x be a symmetrical unlimited variate with probability $F(x)$. Let ξ be defined by

$$\xi = 2n\sqrt{F(x_1)[1 - F(x_n)]}. \tag{3}$$

Then the asymptotic density function $g(\xi)$ and the asymptotic probability $G(\xi)$ of ξ are:

$$g(\xi) = \xi K_0(\xi); \qquad G(\xi) = 1 - \xi K_1(\xi), \tag{4}$$

where K_0 and K_1 are the modified Bessel functions of the second kind and of order zero and one."

Introducing instead of A the parameter u defined by $F(u) = 1 - 1/n$ we have, from (1), approximately for large n

$$(5) \quad F(x_1) = 1/n\left(\frac{u}{-x_1}\right)^k, \quad 1 - F(x_n) = 1/n\left(\frac{u}{x_n}\right)^k, \quad x_1 \leq 0, x_n \geq 0, k > 0.$$

For the variable ξ in Elfving's theorem, we obtain asymptotically

$$(6) \qquad \xi_k/2 = u^k\rho^{-k}.$$

We attach a subscript k to ξ to show its dependence on k. The moments of ξ_k are obtained from a formula given by Watson ([3], p. 388) as

$$(7) \qquad \overline{\xi_k^l} = 2^l\Gamma^2(1 + l/2)$$

and all moments of this variate exist.

3. Estimate of parameters. From N sets, each of n observations, we pick out the largest and the smallest, $X_{n,\nu}$ and $X_{1,\nu}$. We subtract from each observed extreme the central value, m, of the N n observations. If each $x_{n,\nu} = X_{n,\nu} - \mathrm{m} \geq 0$ and $x_{1,\nu} = X_{1,\nu} - \mathrm{m} \leq 0$ the sample size is large enough.

Define $\eta = 1/\rho$ The first two moments of η are, from (7),

$$(8) \qquad \bar{\eta} = \frac{1}{u}\Gamma^2(1 + 1/2k), \qquad \overline{\eta^2} = \frac{1}{u^2}\Gamma^2(1 + 1/k).$$

Elimination of the parameter u from these two equations leads to

$$\frac{\overline{\eta^2}}{\bar{\eta}^2} = \frac{\Gamma^2(1 + 1/k)}{\Gamma^4(1 + 1/2k)}.$$

In terms of the coefficient of variation, V, this equation becomes

$$(9) \qquad \sqrt{1 + V^2} = \Gamma(1 + 1/k)/\Gamma^2(1 + 1/2k).$$

Substituting the value of V computed from the observations, we obtain an estimate of k, and hence can obtain an estimate of u from (8). This procedure is facilitated by Table 1.

4. The distribution of the geometric range. From a practical standpoint the geometric range itself is preferable to its reciprocal since it is easier to interpret and easier to calculate from the observed extremes. We want to establish its distribution $g_1(\rho)$. From the relation (6) of ρ to ξ_k and the knowledge of the distribution (4) of ξ_k we find

$$(10) \qquad G_1(\rho) = 1 - G(\xi_k) = 2u^k\rho^{-k}K_1(2u^k\rho^{-k})$$

and

$$(11) \qquad g_1(\rho) = \frac{2\xi_k\, ku^k}{\rho^{k+1}}K_0(\xi_k) = \frac{4k}{u}\left(\frac{u}{\rho}\right)^{2k+1}K_0\left(\frac{2u^k}{\rho^k}\right).$$

Since tables of these Bessel functions are available [2], the various probabilities and densities may be evaluated.

The simplest way to compare geometric ranges to the theory is the use of a probability paper (Figure 1). For its construction, consider the linear relation

$$\log \rho = \log u + (\log 2)/k - (\log \xi_k)/k \tag{12}$$

obtained from (6). Consequently we plot $-\log \xi_k$ on the abcissa and write the corresponding values $G_1(\rho)$, formula (10), on a horizontal axis. An upper parallel to the abscissa shows the return periods. The observed geometric ranges are plotted on the ordinate in a logarithmic scale. If the theory holds, the observed geometric ranges should be scattered about the straight line (12).

TABLE 1

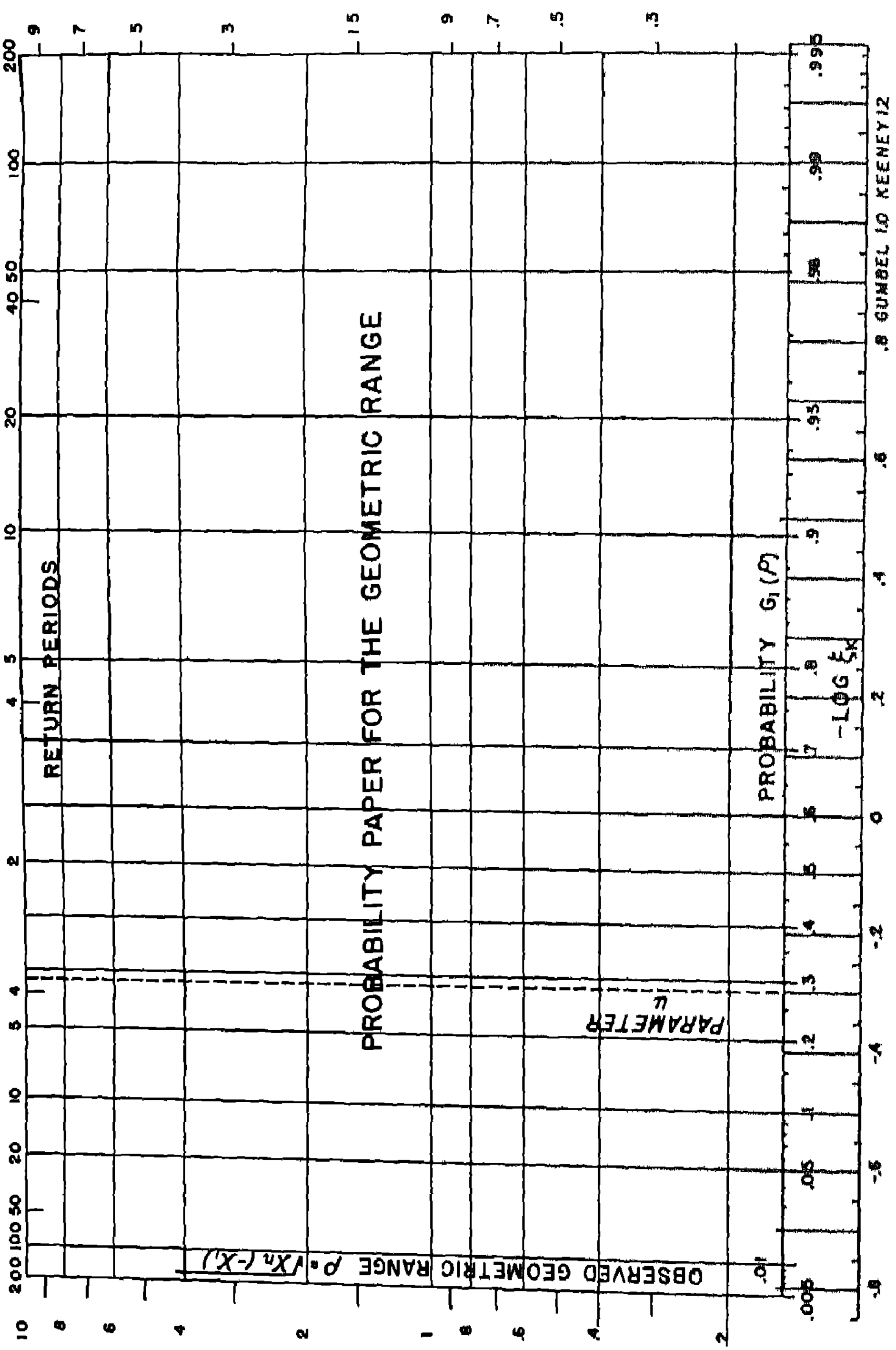
The order k and the variation V of the reciprocal of the geometric range

Reciprocal Order $1/k$	Coefficient of variation V	Reciprocal Order $1/k$	Coefficient of variation V
10	088	.70	556
.12	.104	.80	632
.16	.138	.90	.709
.20	.171	98	772
.30	.251	1 00	788
.40	.332	2 00	1.73
.50	.404	4 00	5.92
.60	.480	6.00	20.0

If less accurate estimates of u and k than those obtainable by the systematic methods (8) and (9), or the probability paper, will suffice, quick estimates can be obtained from the quantiles of the sample of geometric ranges. To the value $\rho = u$ corresponds, according to (6), $\xi_k = 2$ whence, from the tables [2], $G_1(u) = 2K_1(2) = .27973$. From N observed geometric ranges arranged in increasing magnitude we thus may pick out the mth, ρ_m, with the rank $m = .28\,N$ and use it as an estimate $u = \rho_m$. For the medians $\tilde{\xi}_k$ and $\tilde{\rho}$ we get $\tilde{\xi}_k = 1.257$ from the tables, and thus, by (6), $\tilde{\rho}^k = 1.591\,u^k$. This formula provides a quick estimate of k. We pick out the median $\tilde{\rho}$ of the N observed geometric ranges. Since we have an estimate of u, we obtain an estimate of k from

$$\frac{1}{k} = \frac{\log \tilde{\rho} - \log u}{\log 1.591} = 4.960 \log [\tilde{\rho}/\rho_m]. \tag{13}$$

5. Analogy between the geometric range and the range. A study of the various characteristics of the geometric range for distributions of Cauchy's type reveals structural similarities to the range for distributions of the exponential type.


PROBABILITY PAPER FOR THE GEOMETRIC RANGE
RETURN PERIODS
PROBABILITY $G_1(\rho)$
$-\log \xi_K$
PARAMETER
OBSERVED GEOMETRIC RANGE
GUMBEL I.D KEENEY 12

This is not altogether surprising, since (as shown in Table 2) after the appropriate transformations the probabilities of both are identical functions of the respective transformed variates.

Of course the two systems are mutually exclusive: if the observed ranges can be reproduced by the first system we conclude that all moments in the initial distribution exist. If on the other hand, the observed geometric ranges can be represented by the second sytem we conclude that no moments of an order greater than k exist.

TABLE 2
RANGES AND GEOMETRIC RANGES

Type of Initial Distribution	Exponential	Gauchy
Variate	Range	Geometric Range
Definition	$w = x_n + (-x_1)$	$\rho = \sqrt{x_n(-x_1)}$
Transformation	$z = 2\exp\left[-\frac{\alpha}{2}(x_n - x_1 - 2u)\right]$	$\xi_k = 2u^k \rho^{-k}$
Logarithm	$\lg z = \lg 2 - \frac{\alpha}{2}(x_n - x_1 - 2u)$	$\lg \xi_k = \lg 2 - \frac{k}{2}(\lg x_n + \lg(-x_1) - 2\lg u)$
Probability	$G(w) = z K_1(z)$	$G_1(\rho) = \xi_k K_1(\xi_k)$
Distribution	$g(w) = \frac{\alpha z^2}{2} K_0(z)$	$g_1(\rho) = \frac{4k}{u}\left(\frac{\xi_k}{2}\right)^{2k+1} K_0(\xi_k)$
Median	$\tilde{w} = 2u + .9286/\alpha$	$2\lg\tilde{\rho} = 2\lg u + .9286/k$
Mean	$\bar{w} = 2u + 2\gamma/\alpha$	$\lg \overline{\rho^{-1}} = -\lg u + 2\lg\Gamma(1 + \frac{1}{2}k)$

REFERENCES

[1] G. Elfving, "The asymptotical distribution of range in samples from a normal population," *Biometrika*, Vol. 35 (1947).

[2] *Tables of the Bessel-functions*, Vol 6, British Association for the Advancement of Science, Cambridge, 1937.

[3] G. N. Watson, *Theory of Bessel-functions*, Cambridge University Press, 1944.

REMARK ON W. M. KINCAID'S "NOTE ON THE ERROR IN INTERPOLATION OF A FUNCTION OF TWO INDEPENDENT VARIABLES"

By T. N. E. Greville

Federal Security Agency

In a review of Dr. W. M. Kincaid's "Note on the Error in Interpolation of a Function of Two Independent Variables," (*Annals of Math. Stat.*, Vol. 19 (1948),

pp. 85–88) which appeared in *Mathematical Reviews*, Vol. 9 (1948), p. 470, I stated that "a more simple and elegant, and equally general, expression is obtainable by a simple adaptation of formula (41), p. 215, of J. F. Steffensen's book, *Interpolation*."

This statement is not entirely correct and is also misleading in its implications since Dr. Kincaid's expressions are actually more general in certain respects, and simplicity and generality are not the only considerations nor, in this case, the most important ones. In setting up an expression for the remainder in an interpolation formula, the primary objective is to secure an efficient appraisal of the remainder. In this respect, Dr. Kincaid's expressions are superior as they involve only the higher derivatives of the function it is desired to represent, whereas Steffensen's method would always involve a first derivative term in such a way as to prevent any refinement of estimates of the error by introducing additional given values.

REMARK ON MY PAPER "ON A THEOREM OF HSU AND ROBBINS"

By P. Erdös

Syracuse University

Professor Robbins kindly pointed out that in my paper mentioned in the title (*Annals of Math. Stat.*, Vol. 20 (1949), p. 286–291) I have misquoted a statement in the paper of Hsu and Robbins ("Complete Convergence and the Law of Large Numbers" *Proc. Nat. Acad. of Sci.*, Vol. 33 (1947), p. 25–31). I attribute to Hsu and Robbins the conjecture (notations of my paper) that if $\sum_{n=1}^{\infty} M_n < \infty$ then (1) and (2) hold, and proceed to give a counter example. However, the conjecture of Hsu and Robbins is not the above false one but the following: If $\sum_{n=1}^{\infty} M_n < \infty$ and (1) holds then (2) also holds. This conjecture is true and is in fact proved in my paper.

Professor Robbins also points out that a slight modification of my theorem can be stated in a more concise form as follows: Let $X_1, X_2, \cdots$ be a sequence of independent random variables having the same distribution function $F(x)$, and let

$$Y_n = (1/n)\ (X_1 + \cdots + X_n)$$

Then the necessary and sufficient condition that

$$\sum_{n=1}^{\infty} P_r\{|Y_n| > \epsilon\} < \infty, \qquad \text{for every } \epsilon > 0,$$

is that

$$\int_{-\infty}^{\infty} x\, dF(x) = 0, \qquad \int_{-\infty}^{\infty} x^2\, dF(x) < \infty.$$

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the New York meeting of the Institute, December 27–30, 1949)

1. The Asymptotic Distribution of the Extremal Quotient. E. J. Gumbel, New York, and R. D. Keeney, Metropolitan Life Insurance Company, New York.

The extremal quotient is the ratio of the largest to the absolute value of the smallest observation. Its analytical properties for symmetrical, continuous and unlimited distributions are obtained from a study of the auto-quotient defined as the ratio of two non-negative variates with identical distributions. The relation of the two statistics is established by proving that, for sufficiently large samples from an initial distribution with median zero, the largest (or smallest) value may be assumed to be positive (or negative) and that the extremes are independent. The logarithm of the extremal quotient has asymptotically a symmetrical distribution. Its median is unity. As many moments exist for the extremal quotient as moments and reciprocal moments exist simultaneously for the initial variate. For the exponential type of initial distributions, the asymptotic distribution of the extremal quotient can only be expressed by a complicated integral which may be approximated in the interval $\frac{1}{2} < q < 2$ by the logarithmically transformed normal probability function. In this case, no moments exist. For the Cauchy type, the asymptotic distribution of the extremal quotient is very simple. The logarithm of the extremal quotient has the same (logistic) distribution as the midrange for initial distributions of exponential type. For both initial types, the asymptotic distributions of the extremal quotients possess one parameter which may be estimated from the observations

2. A Second Formula for Partial Sums of Hypergeometric Series having the Unit as Fourth Argument. Hermann von Schelling, Naval Medical Research Laboratory, U. S. Submarine Base, New London, Conn.

If the arguments α and β are changed after the summation, published *Ann. Math. Stat.* Vol. *20*, (1949) p. 120, and this method is applied a second time, a new formula results for partial sums of $F(\alpha,\beta,\gamma;1)$. A simple recurrence formula is developed for these partial sums. The new equation is a numerical short cut as it is demonstrated with an example.

3. A Coverage Distribution. Herbert Solomon, Office of Naval Research, Washington, D. C.

Consider a fixed target circle of radius T_R and center at a distance R from an aiming point. Let N circles each of radius W_R be dropped at the aiming point with their centers subject to a bivariate normal distribution with circular symmetry, the common standard deviation denoted by σ. Define γ as the set theoretical sum of the N random circles with the fixed circle and let c be the ratio of γ to the total area of the fixed circle. Then it is desired to find P_{c_0} where

$$P_{c_0} = P\{c \geq c_0 \mid T_R, W_R, R, N\}$$

where T_R, W_R, and R are in σ units. Define $R^* = W_R + aT_R$ where $a = a(c, W_R, T_R)$; $|a| \leq 1$. It is shown that for $N = 1$, the family of curves in the RR^* plane defined by $P_{c_0} =$ constant have a slope, m, given by

$$m = \frac{I_1(RR^*)}{I_0(RR^*)}$$

where I_k is the modified Bessel Function of k^{th} order. In fact as the product

RR^* approaches infinity, m approaches unity. From these results, the contours of equal probability are easily determined. When $N > 1$, overlap considerations make the computation of explicit values for P_{c_0} intractable. However, in this case, upper and lower bounds for P_{c_0} can be obtained.

4. The Problem of the Greater Mean. R. R. Bahadur and Herbert Robbins, University of North Carolina, Chapel Hill.

"Optimum" solutions (in the sense of Wald's theory of statistical decision functions) are obtained for the "problem of the greater mean". Let π_i $(i = 1,2)$ be normal populations with means m_i and common variance σ^2, all unknown, and denote the arbitrary but given set of possible parameter points $\omega = (m_1, m_2 : \sigma)$ by Ω. Suppose that a set of $n_1 + n_2$ independent observations is drawn, n_i from π_i, and let $v = (x_{11}, \cdots, x_{1n_1}; x_{21}, \cdots, x_{2n_2})$ denote the sample point Any measurable function $f(v)$ such that $0 \leq f(v) \leq 1$ is called a decision function. Given a "risk function" $r(f \mid \omega)$ defined for all f and all $\omega \epsilon \Omega$, a decision function $f^*(v)$ is "optimal" if (i) $\sup[r(f^* \mid \omega)] = \inf \sup [r(f \mid \omega)]$, and (ii) no decision function is "uniformly better" than $f^*(v)$. If $f^*(v)$ is the unique (up to sets of measure 0) decision function with property (i), it is "optimum". *Case* 1. Given any decision function $f(v)$ and any $\omega \epsilon \Omega$, let

$$r(f \mid \omega) = \max [m_1, m_2] - m_1 E[f \mid \omega] - m_2 E[1 - f \mid \omega].$$

Let

$$f^\circ(v) = \begin{matrix} 1 \text{ if } \bar{x}_1 > \bar{x}_2 \\ 0 \text{ otherwise} \end{matrix} \qquad \left(\bar{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}\right)$$

It is shown that under certain conditions on Ω, $f^\circ(v)$ is optimum. *Case* 2. Given any decision function which takes on only the values 0 and 1, corresponding to the two decisions "$m_1 \leq m_2$" and "$m_2 \leq m_1$" respectively, and any $\omega \epsilon \Omega$, let

$$r(f \mid \omega) = P(\text{incorrect decision} \mid \omega, f).$$

It is shown that under certain conditions on Ω, $f^\circ(v)$ is optimal. The conditions on Ω are very similar in the two cases, and are likely to be satisfied in most applications. However, it is shown by examples that there exist non-degenerate types of Ω with respect to which decision functions other than $f^\circ(v)$ are *uniformly better* than $f^\circ(v)$. The methods of the paper can be applied to a number of similar problems.

5. Some Extensions of Bayes' Theorem. F. C. Leone, Case Institute of Technology, Cleveland 6, Ohio.

There is some past or *a priori* knowledge about the quality of a population of lots and a sample is taken from a random lot. What can be said about the lot from which this sample is taken? We are incorporating the results of our experiment or sample with the previous knowledge to form a judgment. From the *a priori* distribution and a sample of n with c defectives, say two in twenty-five, we form an *a posteriori* distribution of all two in twenty-five cases From this distribution we can answer questions such as: "What is the *a posteriori* probability that a lot producing a two in twenty-five result should have a proportion of defectives ten per cent or below?" We consider as our *a priori* situation such distributions as the rectangular, triangular, normal, Pearson's Type III and Type I. These extensions are applied to some industrial data. In considering lot quality on one hundred per cent inspection, the *a priori* distributions of these data are mostly U-shaped with some bell-shaped and J-shaped In some cases a Pearson Type I proves to be a good fit for the *a priori* distribution.

6. **On Optimum Selections from Multinormal Populations.** Z W. BIRNBAUM AND D. G. CHAPMAN, University of Washington, Seattle

Let $(X, Y_1, \cdots, Y_n)$ have an $(n + 1)$-dimensional non-singular normal probability density $f(X, Y_1, \cdots, Y_n)$. By "selection" in $(Y_1, \cdots, Y_n)$ we shall understand a measurable function $\varphi(Y_1, \cdots, Y_n)$ such that $0 \leq \varphi \leq 1$ for all $Y_1, \cdots, Y_n$. By a "truncation in $(Y_1, \cdots, Y_n)$ to the set Ω" we understand a selection $\varphi(Y_1, \cdots, Y_n)$ such that $\varphi = 1$ for $(Y_1, \cdots, Y_n)$ in Ω, and $\varphi = 0$ in Ω. A "linear truncation" will be a truncation to a set defined by a condition of the form $\sum_{c=1}^{n} c_i Y_i \geq k$ Using a slight generalization of Neyman-Pearson's fundamental lemma, the following theorems are proven among selections for which the expectation of X, after selection, assumes a fixed value, the one which maximizes the "retained" portion of the universe $\int \cdots \int \varphi(Y_1, \cdots, Y_n) f(X, Y_1, \cdots Y_n) dX dY_1 \cdots dY_n$ is a linear truncation. Among all the selections for which a given quantile of X, after selection, assumes a fixed value, the one which maximizes the retained portion of the universe is a linear truncation. (Research under the sponsorship of the Office of Naval Research).

7. **Simple Regression Analysis with Autocorrelated Disturbances.** HOWARD L. JONES, Illinois Bell Telephone Company, Chicago.

When the disturbances in a regression equation are connected by a linear difference equation, the parameters of both equations can be estimated simultaneously by maximizing a function that describes the joint probability of the disturbances or a linear function thereof This note discusses a simple example.

8. **A Test of Klein's Model III for Changes of Structure.** A. W. MARSHALL, The Rand Corporation, Santa Monica, Calif.

This paper suggests a test of equations from linear stochastic equation systems on the basis of observations not included in the original computation period. Rejection regions of approximately the right size (asymptotically correct) are constructed and the use of naive economic models as an auxillary test are suggested The procedure is applied to Klein's Model III, the results are tabulated and discussed.

9. **An Application of the Theory of Extreme Values to Economic Problems.** S. B. LITTAUER, Columbia University, AND E. J. GUMBEL, New York.

Most studies of economic time series have been concerned with establishing regularities of behavior, often by analogy with mechanical systems Much as regularity in economic phenomena is desirable, such evidence as has been available leaves the reality of this sought for regularity considerably in doubt It seems more fruitful rather to ask the question, "What is the pattern of the non-regularity" and if reasonably answered, to offer some verifiable form of explanation therefor. It seems further desirable that any attempt at "scientific" explanation of economic phenomena be fortified by evidence of statistical stability supported by criteria such as were established by Shewhart for the control of quality of manufactured product In the present instance certain concepts of experimental inference, which seem natural therefor, are employed in order to give some general and plausible unity to the behavior of economic time series.

Following upon the postulates of the theory presented here, the appropriate formal development employs concepts of statistical quality control and of the statistical theory of extreme values Within this theory the importance of the absence of statistical stability

is emphasized, and the relevance of the use of concepts in extreme values is made evident. By introducing a superuniverse, peaks and troughs are random expressions of a super chance-"cause" system The use of these statistical concepts is not motivated by mere analogy but rather as the natural means for explanation of the phenomena studied.

A number of examples of the application of these statistical methods to selected series are offered as evidence of the workability of the theory here presented. The extremes of the Dow-Jones index of selected industrials show that the 1928 value was completely outside the previous levels and should not have been considered as a "stable high plateau basic for perpetual prosperity". Instead this should have suggested the imminent breakdown. The validity of the application of the theory of extreme values to these phenomena is not so strongly substantiated as are the many applications that have been made of them to flood frequencies, wind velocities, extreme temperatures, breaking strengths and other natural phenomena. Nevertheless the results here obtained are highly suggestive of a tenable economic hypothesis.

10. Bias Due to the Omission of Independent Variables in Ordinary Multiple Regression Analysis. (Preliminary Report). T. A. BANCROFT, Iowa State College, Ames.

Given n observations of the dependent variate y and the independent variates x_1, $x_2, \cdots, x_k, \cdots, x_r$, $k < r$, all variates measured from their respective sample means, and we have calculated the ordinary regression of y on the first k variates and y on all r variates. We define ordinary multiple regression as the single-equation approach, error only in y which is assumed normally and independently distributed with zero mean and variance σ^2, the x_i being fixed from sample to sample.

In order to determine whether to omit or retain the last $(r - k)$ independent variates we formulate a rule of procedure: calculate Snedecor's $F =$

$$\frac{\text{Reduction in } Sy^2 \text{ due to } (r-k) \text{ variates}/(r-k)}{\text{Error mean square after fitting all } r \text{ variates}}$$

If F is non-significant at some assigned significance level α, we pool the sums of squares and degrees of freedom, involved in the numerator and denominator of F, to obtain an estimate of the error σ^2, and fit y on the first k variates only. If F is significant at the assigned significance level we use the denominator only in F for our estimate of σ^2 and hence fit y on all r variates.

The object of this investigation is to determine the bias in our estimate s^* of σ^2, if we follow such a rule of procedure. The bias turns out to be

$$\frac{2\sigma^2\lambda}{n_1+n_2} + \sigma^2 e^{-\lambda} \sum_{i=0}^{\infty} \left[I_{x_0}\left(\frac{n_2}{2}+1, \frac{n_1}{2}+i\right) - I_{x_0}\left(\frac{n_2}{2}, \frac{n_1}{2}+i\right) \frac{-2i}{n_1+n_2} I_{x_0}\left(\frac{n_2}{2}, \frac{n_1}{2}+i\right)\right] \frac{\lambda^i}{i!},$$

where

$$x_0 = \frac{n_2}{n_2 + n_1\alpha}, \qquad \lambda = \frac{\sum_{i=k+1}^{r} (\beta_i')^2}{2\sigma^2},$$

n_1 and n_2 are the respective degrees of freedom for the numerator and denominator of F, and $\sum_{i=k+1}^{r} (\beta_i')^2$ is a function of the population regression coefficients $\beta_{k+1}, \cdots, \beta_r$. The bias is discussed for selected values of the parameters involved.

11. Estimating Parameters of Pearson Type III Populations From Truncated Samples. A. C. COHEN, JR., The University of Georgia, Athens.

The method of moments is employed with 'single' truncated random samples (1) to estimate the mean, μ, and the standard deviation, σ, of a Pearson Type III population when α_3 is known and (2) to estimate μ, σ, and α_3 when only the form of the distribution is known in advance. No information is assumed to be available about the number of variates in the omitted portion of the sample. The results obtained can be readily applied to practical problems with the aid of "Salvosa's Tables of Pearson's Type III Function." An illustrative example is included in the paper.

12. The Cyclical Normal Distribution. E. J. GUMBEL, New York.

The usual normal distribution becomes invalid for variates, like an angle, lying on the circumference of a circle The distribution of such variates was established by R. von Mises by the same methods as used for the classical derivation. The cyclical normal distribution is symmetrical about a mode and antimode. The probability function is proportional to an incomplete Bessel function of the first kind and of order zero for an imaginary argument, and contains two parameters, the direction of the resultant vector and a parameter k linked to the absolute amount of the vector. The parameters may be estimated by the method of maximum likelihood. For $k = 0$, the distribution degenerates into a uniform cyclical distribution. If k is of the order 3, the distribution approaches the linear normal one, k being the reciprocal of the variance. With increasing values of k, the distribution looses its cyclical character and becomes concentrated in a narrow strip This distribution holds for symmetrical unimodal values varying according to pure chance about a unique mode in a closed space (as the angles of the wind directions) or a closed time, and gives a theoretical model for the variations of temperatures, pressures, rainfalls, storms, discharges, floods, death- and birth rates over the year, and earth quakes over the day. The comparison between theory and observations in plotting the square roots of the frequency on polar coordinate paper provides a statistical criterion for the regularity of cyclical phenomena (Work done in part under contract W 44/109/QM/2202 with the Research and Development Branch, Office of the Quartermaster General)

13. Treatment of Attenuation Problems by Random Sampling. H. KAHN AND T. HARRIS, The Rand Corporation, Santa Monica, Calif.

Exact analytical calculations of the transmission of energy by particles through shields are difficult; to avoid them random sampling methods may be resorted to. The straightforward procedure of simulating life histories of particles, using random number tables, may be used for thin shields, but in the case of thick shields with tremendous attenuations, tremendous numbers of particles would be required. In order to obtain reasonably small standard errors, using reasonable numbers of simulated life histories, it is necessary to modify the original problem to one having a lower attenuation factor, the solution bearing a known relation to the solution of the original problem. Alternatively, this may often be regarded as an application of well known statistical sampling procedures, such as representative sampling or importance sampling. Various special procedures can be devised. One of the first was the splitting technique due to J. v. Neumann. Among others may be mentioned the exponential transformation, a simple analytic transformation of the original problem into one having a much lower attenuation factor.

14. On the Existence of Nearly Locally Best Unbiased Estimates. HERMAN RUBIN, Stanford University, Stanford, Calif.

For any family $\mathcal{F}$ of distributions, and any distribution F_0 of $\mathcal{F}$, there exists a bilinear function K whose arguments are all parameters defined for all distributions of $\mathcal{F}$ and for

which there exist unbiased estimates which have finite variance if F_0 is the true distribution, and which has the following properties: (1) If θ is any parameter in the domain of K, and t is any unbiased estimate of θ, then $var(t \mid F_0) \geq K(\theta, \theta)$. (2) This result is best possible, i. e., for any θ there is an unbiased estimate t of θ whose variance differs from $K(\theta, \theta)$ by less than any preassigned amount

15. The Experimental Evaluation of Multiple Definite Integrals. George W. Taylor, U. S. Army Electronics Laboratory, San Diego, Calif.

When one is forming an estimate of the total, or mean value, of some quantity, sampling at carefully selected points will frequently be preferable to employing a method which involves randomization. The estimation of the total volume of water in a given lake or the amount of energy being released in a given time and space, are examples of problems where specified points for sampling should result in a reduction in the error of estimate. These and similar problems lead naturally to numerical integration methods. In the case of single integrals, Gauss' and Tchebychef's formulae yield maximum efficiency with respect to controlling the polynomial error and statistical error respectively, but often the Newton-Cotes formulae can be applied more conveniently.

For the evaluation of double integrals, an eight point and a thirteen point formula for fifth degree accuracy and a twelve point and a twenty-one point formula for seventh degree accuracy have been developed for integrating over a rectangle and similar formulae have been developed for integrating over areas bounded by a parabola and a straight line or by two parabolas. The following system of equations is employed in developing these formulae:

$$\sum_{\alpha=1}^{m} R_\alpha x_\alpha^i y_\alpha^j = C_{ij}, \quad \text{for all } i, j \text{ for which } i + j \leq 2n,$$

$$\text{and where } C_{ij} = \frac{a^i b^j}{(i+1)(j+1)} \text{ for both } i \text{ and } j \text{ even,}$$

$$= 0 \text{ otherwise.}$$

Formulae for the numerical evaluation of triple integrals taken over a rectangular parallelopiped are developed, including a twenty-one point formula with fifth degree accuracy. It is shown that comparable formulae can be developed for integrating functions of more than three variables and a $2n + 1$ point formula with third degree accuracy for integrating a function of n variables over a rectangular n-space is obtained.

16. Tests of Fit of a Cumulative Distribution Function over Partial Range of Sample Data. Bradford F. Kimball, New York State Dept. of Public Service, New York.

Case 1 Sample data are completely ordered over range tested.

Let the $n + 1$ true frequency differences associated with an ordered random sample of n values of x be denoted by u_i. The *cdf* of a theoretical test function based on m of the above frequency differences is identified and methods of approximating it are discussed.

Case 2. Sample data in k ordered groups over range tested.

Let $\Delta_i F$ denote the true frequency differences over the k sample intervals to be covered by the test. Let m_i denote the number of unit frequency differences u_i covered by the ith interval. Define M and W by

$$M + 1 = \sum_k m_i, \qquad M \leqq n,$$

$$W = \sum_k \Delta_i F, \qquad W \leqq 1.$$

A theoretical function Z is defined by

$$Z = \frac{(M+1)(M+2)}{k-1} \sum_k \frac{[\Delta_i F - m_i W/(M+1)]^2}{m_i}.$$

Set

$$Y = Z/W^2.$$

The *cdf* of Y is identified and methods of approximation to it are discussed

Applications to testing agreement of sample with hypothetical *cdf* of universe are considered for both cases in some detail.

17. Large Sample Tests for Comparing Percentage Points of Two Arbitrary Continuous Populations. A. W. MARSHALL AND J. E. WALSH, The Rand Corporation Santa Monica, Calif.

Let us consider two continuous populations, the first with density function $f(x)$ and $100\alpha\%$ point θ_α, the second with density function $g(x)$ and $100\beta\%$ point ϕ_β These two populations are arbitrary except that $f(\theta_\alpha) \neq 0$, $g(\phi_\beta) \neq 0$ and both $f'(\theta_\alpha)$, $g'(\phi_\beta)$ exist and are continuous in the vicinity of the specified points This paper presents significance tests for $\theta_\alpha - \phi_\beta$ which are based on large samples from these populations The exact significance level of a test is not known but its value is bounded within reasonably close limits (asymptotically). Efficiency properties of these tests (compared to the corresponding noncentral t-tests) are investigated for the case in which both populations are normal and the ratio of variances is known Results are also derived for simultaneously testing $\theta_\alpha - \phi_\beta$ and $f(\theta_\alpha)/g(\phi_\beta)$ These tests have known significance levels (asymptotically). A particular application of tests of this type occurs when it is desired to test whether two samples came from the same population and agreement of the two populations in a specified region is to be emphasized. For this special case, the significance levels of the resulting tests are reasonably accurate for moderate as well as large sized samples

18. On the Distribution of Wald's Classification Statistic. H. L. HARTER, Michigan State College, East Lansing.

A study is made of the distribution of the classification statistic introduced by Wald. The exact distribution of V in the univariate case, as obtained by the use of characteristic functions and contour integration, is given for both degenerate and non-degenerate cases. The problem of classifying an individual into one or the other of two populations, using the statistic V, is discussed. In the multivariate case, examples are given of the distribution of an approximation to V suggested by Wald The procedure here consists integrating out two variables from the joint distribution of three variables to find the distribution of the third. Four cases arise, depending upon whether the sample size and the number of variates are even or odd. Since this approximation is valid only for large samples, an attempt is made to find an approximation which is asymptotically equivalent to it as the sample size increases, but which is valid also for small samples. Results are given for a sampling experiment performed to determine an empirical distribution of V for a specific small sampling case, using a population of 10,000 pieces modeled after Shewhart's normal bowl. Obstacles in the path of practical applications are discussed

19. Analysis of Extreme Values. W. J. DIXON, University of Oregon, Eugene.

Consider a population $N(\mu, \sigma^2)$ contaminated by introducing a certain proportion of values from a population $N(\mu + \lambda\sigma, \sigma^2)$ or $N(\mu, \lambda^2\sigma^2)$ The performance of various statistics for discovering these contaminators is assessed by sampling methods for samples of size 5 and 15 (This research was sponsored by the Office of Naval Research)

20. **A Note On The Variance Of Truncated Normal Distributions.** A. C. COHEN, JR., The University of Georgia, Athens.

Formulas are derived whereby the variance of truncated normal distributions can readily be computed with the aid of an ordinary table of areas and ordinates of the normal frequency function These results are applicable to certain tolerance problems involved in Statistical Quality Control Their use will enable one to make computations required in solving such problems without resorting to Karl Pearson's relatively inaccessible tables of "Values of the Incomplete Normal Moment Functions".

21. **Some Estimates and Tests Based on the r Smallest Values in a Sample** (By Title). J. E. WALSH, The Rand Corporation, Santa Monica, Calif.

Let us consider a situation where only the r smallest values of sample of size n are available. This paper investigates the case where n is large and r is of the form $pn + 0(\sqrt{n})$. Properties of some well known estimates and tests of the $100p\%$ population point (based on statistics of the type used for the sign test) are investigated. If the sample is from a normal population, these nonparametric results have high efficiencies for small values of p (at least 95% if $p \leq 1/10$). The other investigations are restricted to the case of a normal population. Asymptotically "best" estimates and tests of the population percentage points are derived for the case where the population variance is known. If the population variance is unknown, asymptotically most efficient estimates and tests can be obtained for the smaller population percentage points by suitable choices of p and $0(\sqrt{n})$. The results of the paper have application in the field of life testing. There the r smallest sample values can be obtained without the necessity of obtaining the remaining sample values By starting with a larger number of units but stopping the experiment when only a small percentage have "died", it is often possible to obtain the same amount of "information" with a substantial saving in cost and time over that required by starting with a smaller number of units but continuing until all have "died".

22. **Some Comments on the Efficiency of Significance Tests** (By Title) J. E. WALSH, The Rand Corporation, Santa Monica, Calif.

A method sometimes used to measure the efficiency of a significance test consists in associating a statistic with the test and defining the efficiency of the test to be the efficiency of this statistic considered as an estimate. This paper investigates the power function implications of this method of defining the efficiency of a test. Examples are presented which show that an estimate efficiency of $100E\%$ does not necessarily imply that the corresponding most powerful test based on $100E\%$ as many sample values has approximately the same power function as the given test (for the admissible set of alternative hypotheses). In several of the examples it was found that estimate efficiency makes no allowance for the effect of significance level while the relationship between the power functions of the given test and the corresponding most powerful test changes noticeably with respect to significance level. Some of these examples are non-asymptotic while others are asymptotic. However, results are obtained for the asymptotic case which indicate that this equality of power functions does hold for a rather broad class of significance tests if the pertinent statistics have distributions which are asymptotically normal.

23 **Application of Sequential Sampling Method to Check the Accuracy of a Perpetual Inventory Record.** JOSEPH B. JEMING, New York.

The problem of checking the continuing property records of a large utility company is handled by an application of the sequential sampling method as developed by the Statistical Research Group,

Columbia University. Without the application of a sampling procedure the problem can only be solved either by a complete physical inventory which is very costly, or by a cycle check which takes many years to complete. By use of the sequential sampling method, results of desired accuracy are obtained quickly and at very low cost since an extremely small percentage of field inspection for the mass property accounts of any large utility produces satisfactory conclusions.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest.

Personal Items

Dr. Ralph A. Bradley accepted an appointment as Assistant Professor in the Mathematics Department of McGill University, Montreal, Canada after receiving his Ph.D. in mathematical statistics at the University of North Carolina in June, 1949.

Mr. Fred J. Clark, Jr. received his master of science degree in mathematics from the University of Illinois in August, 1949 and is now employed by the University of California at the Sandia Laboratory in Albuquerque, New Mexico.

Professor J. L. Doob is on leave from the University of Illinois to teach at Cornell University for the academic year 1949–1950.

Mark W. Eudey obtained his Ph.D. degree in statistics at the University of California, Berkeley, and is now Vice President of California Municipal Statistics, Inc.

Dr. Joseph L. Hodges, Jr. has been promoted to Assistant Professor and Research Associate at the Statistical Laboratory, University of California, Berkeley.

Professor Paul Horst, formerly of the Department of Psychology, University of Washington, is now Director of Research at the Educational Testing Service, Princeton, New Jersey.

Dr. Fred C. Leone, formerly an Instructor and a Research Fellow at Purdue University, has been appointed Instructor in the Mathematics Department and Director of the Statistical Laboratory at the Case Institute of Technology.

Mr. Fred W. Lott, who has been studying at the University of Michigan for his Ph.D., has accepted an assistant professorship at Iowa State Teachers College, Cedar Falls, Iowa.

Dr. Francis McIntyre has resigned as Director of Export Control, Office of International Trade, U. S. Department of Commerce, Washington, D. C. to accept a post as Director of Economic Research, California Texas Oil Co., 551 Fifth Avenue, New York, New York.

Mr. R. B. Murphy, who has been a graduate student at Princeton University has accepted an instructorship in the Mathematics Department of Carnegie Institute of Technology.

Professor Jerzy Neyman, Director of the Statistical Laboratory, University of California at Berkeley, will be on sabbatical leave for the Spring Semester, 1950.

Mr. Monroe L. Norden, formerly of the Glenn L. Martin Co., is now a Mathematical Statistician with the Operations Research Office, Johns Hopkins University, Ft. Lesley, J. McNair, Washington 25, D. C.

Mr. D. Martin Sandelius, formerly a Research Assistant in the Institute of Statistics, Uppsala, Sweden, has been appointed Lecturer in the Mathematics Department, University of Washington, Seattle, for the academic year 1949–1950.

After completing his graduate work at Ohio State University, Dr. William J. Schull accepted a position with the Atomic Bomb Casualty Commission. He is now in Japan as a geneticist working on follow-up studies at Hiroshima.

Miss Elizabeth L. Scott obtained her Ph.D. degree in statistics at the University of California, Berkeley and was promoted to Lecturer and Research Associate at the Statistical Laboratory.

Miss Ester Seiden obtained her Ph.D. degree at the University of California, Berkeley and was promoted to Lecturer and Research Associate at the Statistical Laboratory.

Mr. Iriving H. Siegel is on leave from his position as Chief Economist at the Veterans Administration until June 30, 1950, to serve as Lecturer in Political Economy at the Johns Hopkins University and as a member of the Johns Hopkins University Operations Research Office staff.

Dr. Charles M. Stein, Assistant Professor and Research Associate at the Statistical Laboratory, University of California, Berkeley, will be on leave for the academic year 1949–1950 and will be working in Paris as a National Research Fellow.

Alfred James Lotka

Alfred James Lotka, a Fellow of the Institute, died in Red Bank, New Jersey, on December 5, 1949 He was born of American parents in Poland, March 2, 1880, and had his early schooling in France. His academic training was received at Birmingham, England (B.Sc., 1901, and D.Sc., 1912), Cornell (M.A., 1909), and Johns Hopkins (1922–1924). Dr. Lotka came to the Statistical Bureau of the Metropolitan Life Insurance Company in 1924 and retired as Assistant Statistician in 1947. His major contributions were his highly original work on the mathematical theory of evolution, on the mathematical analysis of population, and on the theory of self-renewing aggregates. Altogether, Dr. Lotka had almost 100 papers in these fields in technical and scientific journals, both here and abroad. The essentials of his work are summarized in his books, "The Elements of Human Biology" and "Theorie analytique des associations biologiques." He was, in addition, a joint author on several books in the field of public health.

Dr. Lotka was a past president of the American Statistical Association and of the Population Association of America. He had recently been active as American Vice-President of the International Union for the Study of Population.

Statistical Summer Session in Berkeley, Calif.

Following the established pattern, there will be held this year a Statistical Summer Session at the University of California, Berkeley. The faculty will include William G Cochran of Johns Hopkins University, Benjamin Epstein of Wayne University, Erich L. Lehmann of the University of California, Paul Lévy

of the Ecole Polytechnique, Paris, France and Gottfried E. Noether of New York University.

Courses will be offered on both the graduate and the undergraduate levels. The graduate courses, all given during the First Summer Session, June 19 to July 29, are meant primarily for students who either have already obtained their Ph.D. degree or are working toward it. No specific prerequisites to graduate courses will be required. The graduate program includes (i) a course on design of experiments and a seminar on analysis of variance by W G. Cochran, (ii) a course on theory of estimation by E. L. Lehmann, and (iii) a course and a seminar on random variables and random functions by Paul Lévy.

Inquiries should be addressed to the Office of the Summer Sessions, 1A Administration Building, University of California, Berkeley 4, California.

At a meeting of its Executive Council, AAPOR has laid plans for its 1950 meetings to be held jointly with the World Association for Public Opinion Research (WAPOR) at Lake Forest College, near Chicago, June 16 to 20.

The program which is now being planned will be designed to fit the needs of the Association's membership, which is composed of leaders in both the academic and commercial fields.

The Council of the Institute of Mathematical Statistics requested Professor Harold Hotelling to communicate to Professor S. S. Wilks its appreciation of his editorship of the *Annals* during the years 1938 to 1949. On the recommendation of the Council Professor Hotelling's letter is reproduced below.

January 6, 1950

Professor Samuel S. Wilks
Fine Hall
Princeton, New Jersey
Dear Professor Wilks:

In behalf of the Council of the Institute of Mathematical Statistics and by its direction, I write to express the appreciation we all feel for the splendid efforts which you have expended so freely upon the *Annals of Mathematical Statistics*, and which have been so conspicuously successful in establishing it as a sound and reputable journal. The years of your editorship are memorable ones for the history of statistics, and your contribution to making them so is of first importance.

Very sincerely,
Harold Hotelling

New Members

The following persons have been elected to membership in the Institute

(August 23, 1949 to November 30, 1949)

Anderson, Oskar, Ph.D. (Kiel) Professor, University of Munich, *Konigin-Strasse 69, Munich (Munchen), Germany*

Puente Arroyo, Felix Jorge, CPA, (Univ. Nal. Litoral) Professor titular Mathematics, *Italia 1550, Rosario, Republica Argentina.*

Arvanitis, Ernest A., A.B. (Boston Univ.) Student at Columbia University, *43–18 40th Street, Sunnyside, L. I., New York*

Bhatt, Narbheshanker M., Ph.D. (Edinburgh Univ.) Professor of Statistics, Commerce College, Behind Raopura Tower, Baroda, India

Bose, Raj Chandra, D. Litt (Calcutta Univ.) Professor of Mathematical Statistics, University of North Carolina, *110 Noble Street, Chapel Hill, North Carolina.*

Carreiro, Oscar Edlwaldo Porto, Civil Engineer (Univ. of Brazil) Professor da Faculadade de Ciencias Economicas, Avenida Sao Sebastiao 266, Sao Paulo, Brazil.

Crump, Phelps P., B.S. (Iowa State) Graduate Student and Research Assistant, *Box 5457, State College Station, Raleigh, North Carolina*

Davis, Richard L., B.S. (North Carolina State) Sales Engineer, *Box 304, Charlotte, North Carolina*

Dickman, Sidney, A.B. (Brooklyn College) Graduate Student at Columbia University, *2823 West 25th Street, Brooklyn 24, New York*

Fitzgerald, Rev. John F., S J., M.S. (Univ. of Detroit) Assistant Professor of Physics and Mathematics, College of the Holy Cross, Worcester 3, Massachusetts.

Godsey, Ellis B., B.S. (Indiana Univ.) Analytical Statistician, Army Chemical Corps, *1716 Pin Oak Road, Baltimore 4, Maryland*

Ghurye, S. G. M.Sc. (Univ of Bombay) Student and assistant, Department of Mathematical Statistics, *c/o The Institute of Statistics, Phillips Hall, Chapel Hill, North Carolina*

Gutt, Paul, M.S. (Univ of Chicago) Ordnance Research #1, Mathematician, *6421 S. Ellis, Chicago, Illinois*

Harman, Harry H., S.M. (Univ of Chicago) Chief, Statistical Research and Analysis Unit, Personnel Research Section, AGO, Dept. of the Army, *4111 Maryland Ave. (Brookmont), Washington 16, D. C.*

Henderson, Charles R., Ph.D. (Iowa State) Associate Professor, Animal Husbandry Department, Cornell University, Ithaca, New York.

Harter, Harman Leon, Ph.D. (Purdue Univ.) Assistant Professor of Mathematics, Michigan State College, East Lansing, Michigan.

Hoffman, William Charles, M A (Univ. of Calif. at Los Angeles) Graduate Assistant, Department of Mathematics, Cornell University, Ithaca, New York.

Hydeman, William Robert, M.A. (Syracuse Univ.) Mathematician, U. S. Navy Department, *3810—39th Street, N.W., Washington 16, D. C.*

Kellerer, Hans, Ph.D Referent, Bayerisches Statistisches Landesamt, Munchen 8, Rosenheimerstr 130, Germany.

Kramer, Kenneth H., M S. (Carnegie Inst. of Tech.) Teaching Assistant at Carnegie Institute of Technology, *279 Seneca Street, Turtle Creek, Pennsylvania*

Lieberman, Gerald J., M.A (Columbia Univ.) Engineer and Mathematical Statistician, Statistical Engineering Laboratory, National Bureau of Standards, Washington 25, D. C.

Lindley, Dennis V., M.A (Cantab) University Demonstrator in Mathematics, Statistical Laboratory, St Andrews Hill, Cambridge, England.

Malan, A. P., M.Sc. (South Africa) Professor, U.C.O.F.S., Bloemfontein, South Africa.

Rasch, G., Ph D. (Copenhagen) Chief of Statistical Department, State Serum Institute, Copenhagen, Denmark

Recao, Manuel Felipe, B.A. (Univ. Venezuela) Director General de Estadistica, Minesterio de Fomento, Professor of Mathematics, Facultad Ciencias Economics, Central University, Calle Real Chacao, Quinta "La Paz," Chacao, Estado Miranda, Venezuela.

Riggs, Charles L., Ph D (Univ. of Kentucky) Assistant Professor of Mathematics, Department of Mathematics, Kent State University, Kent, Ohio.

Saxer, Walter, Ph.D Professor a.d. Eidg. Techn Hochschule, Zurich, Goldbach-Kusnacht, Switzerland.

Scobert, Whitney, M.S. (Univ. of Oregon) Associate Professor of Mathematics, Mathematics Department, Idaho State College, Pocatello, Idaho.

Serfling, Robert E., Ph.D. (Univ. of Mich.) Senior Scientist, Officer in Charge, Statistical Branch, Epidemiology Division, Communicable Disease Center, U. S. Public Health Service, Atlanta, Georgia

Steyn, Hendrik S., Ph.D. (Univ. of Edinburgh) Lecturer in Statistics, University of Pretoria, *306 Fourth Private Avenue, Villieria, Pretoria, South Africa*

Zacharias, William B., A.M. (Univ of Pennsylvania) Instructor in Mathematics, Temple University, *1529—67th Avenue, Philadelphia 26, Pennsylvania*

Zeigler, R. K., Ph.D. (Univ. of Iowa) Associate Professor of Mathematics, Mathematics Department, Bradley University, Peoria 5, Illinois.

REPORT OF THE NEW YORK MEETING OF THE INSTITUTE

The twelfth Annual Meeting of the Institute of Mathematical Statistics was held in New York City on December 27–30, 1949. Headquarters were at the Biltmore Hotel where most of the sessions were held; one or more of the sessions were held at the Hotel Commodore, the McAlpin Hotel, and the Governor Clinton Hotel. The meeting was held in conjunction with the Annual Meeting of the American Statistical Association, the American Association for the Advancement of Science, the American Mathematical Society, the Econometric Society, the Psychometric Society, the Mathematical Association of America, the Association for Computing Machinery, and the American Psychological Association. The following 214 members of the Institute attended:

F. S. Acton, P. H. Anderson, R L. Anderson, T. W Anderson, H. E. Arnold, K J Arnold Max Astrachan, R R. Bahadur, E W. Bailey, T. A. Bancroft, W. D. Baten, E. E. Blanche, C. I. Bliss, R. C. Bose, A. H Bowker, R A. Bradley, Dorothy Brady, A E Brandt, I. D. J. Bross, T H. Brown, O. P. Bruno, P. T. Bruyere, R. W Burgess, J. M. Cameron, B H. Camp, E W. Cannon, S. D. Canter, Bernard Carol, O. S. Carpenter, Maria Castellani, Jack Chassan, Randolph Church, Edmund Churchill, W. G Cochran, A. C. Cohen, Jr , R. H. Cole, E. P Coleman, F. G. Cornell, Jerome Cornfield, C. C. Craig, M T. Crapsey, J F Daly, D. A Darling, Besse B. Day, F. R. Del Priore, W. E Deming, Philip Desind, W. J. Dixon, C. W. Dunnett, Solomon Dutka, P. S. Dwyer, Benjamin Epstein, W. D Evans, W T Federer, William Feller, J W. Fertig, Leon Festinger, C H. Fischer, J. C Flanagan, M. M. Flood, L. R. Frankel, N. M. Franklin, H A. Freeman, Bernard Friedman, Melitta L Garbuny, E. F. Gardner, M. A. Geisler, H. H. Germond, Leon Gilford, Abraham Golub, William Gomberg, C. H. Graves, S. W. Greenhouse, J. A Greenwood, Evelyn S. Grossman, H. T. Guard, Carl Hammer, E. C Hammond, H. H Harman, T E Harris, Boyd Harshbarger, H. L. Harter, W. A. Hendricks, L. H. Herbach, J. L. Hodges, Jr., Wassily Hoeffding, Helen M. Humes, Harold Hotelling, Cuthbert Hurd, H. M. Hughes, W. R. Hydeman, S M. Ikhtiar-ul-Mulk, S. I Isaacson, Marcus Jacobs, W W. Jacobs, J. E. Jackson, Carol M. Jaeger, J. B. Jeming, R. J. Jessen, H. L. Jones, Alice S. Kaitz, W. C. Kalinowski, Leo Katz, R D. Keeney, B. F. Kimball, Leslie Kish, Lila F. Knudsen, Paul Koditschek, C. F. Kossack, K H. Kramer, R. R Kuebler, Jr., S. M. Kwerel, R. B. Ladd, Marguerite Lehr, F C. Leone, Joseph Lev, Howard Levene, G. J. Lieberman, Julius Lieblein, S. B Littauer, Simon Lopata, Irving Lorge, E. D Lowry, L. H. Madow, W. G. Madow, Benjamin Malzberg, Joseph Mandelson, E. S. Marks, Margaret P. Martin, J W. Mauchly, P. J. McCarthy, Margaret Merrell,

Albert Mindlin, P D. Minton, Robert Mirsky, A. M. Mood, Doris N. Morris, R H. Morris, Dorothy J. Morrow, J. W. Morse, J E Morton, Judith Moss, R. G Moss, Frederick Mosteller, C. M Mottley, Hugo Muench, L. F Nanni, Doris Newman, G. E. Noether, M. L. Norden, J. A. Norton, Jr., H. W Norton, E. G Olds, P S. Olmstead, A. L. O'Toole, W. R. Pabst, Jr., R. E. Patton, Katherine Pease, G. W. Petrie, B. E. Phillips, E W. Pike, Aditya Prakash, Frank Proschan, J. E. Raup, L. J. Reed, J. S. Rhodes, P. R. Rider, H. G. Romig, Norman Rudy, Marion M. Sandomire, F. E. Satterthwaite, Mary Ann Savas, M. A. Schneiderman, Samuel Schweid, O A. Shaw, G. D. Shellard W. A. Shewhart, S. S. Shrikhande, Harry Shulman, I H. Siegel, Rosedith Sitgreaves, G. W. Snedecor, Herbert Solomon, D. E. South, Mortimer Spiegelman, R. G D. Steel, J R. Steen, Arthur Stein, Joseph Steinberg, F. F. Stephan, A. I. Sternhell, J. S. Stock, J. G. Strieby, J. V. Sturtevant, W. R. Thompson, L. J. Tick, Gerhard Tintner, M M. Torrey, J W. Tukey, G. W. Tyler, S A. Tyler, Uttam Chand, D. F Votaw, Jr., Helen M. Walker, W. A. Wallis, Samuel Weiss, E. L. Welker, D. R Whitney, Frank Wilcoxon, R. I. Wilkinson, S S. Wilks, C. P Winsor, M. A. Woodbury, Holbrook Working

The opening session on Tuesday, December 27, 9 A. M., held jointly with the American Statistical Association and the American Mathematical Society, was devoted to *Operations Research*, with Professor J. Steinhardt, Operations Evaluation Group, Massachusetts Institute of Technology presiding. The following papers were presented:

1. *Topics on the Methodology of Operations Research.* B. O Koopman, Columbia University.
2. *Some Applications of the Mathematical Theory of Games.* G. E. Kimball, Columbia University.
3. *Theory of Games.* L. Gillman, Operations Evaluation Group, Massachusetts Institute of Technology
4. *Development of Theories of Action.* Ellis Johnson, Operations Research Office. The Johns Hopkins University.
5. *Some Industrial Applications of Operations Research.* A. A. Brown, Operations Evaluation Group, Massachusetts Institute of Technology.

At the second session, held jointly with the American Statistical Association, at 2:30 P M. on the opening day, Professor M. Loeve, University of California, gave a special invited address entitled, *Fundamental Limit Theorems in Probability.* The discussion was presented by Professor Will Feller of Cornell University and Professor H. E. Robbins of the University of North Carolina. Professor Abraham Wald of Columbia University served as chairman.

The first contributed papers session was held on the same day at 4:00 P. M., with Professor W. D. Baten of Michigan State College and Michigan Agricultural Experiment Station as chairman. The following papers were presented:

1. *The Asymptotic Distribution of the Extremal Quotient.* E. J. Gumbel, New York, and R. D Keeney, Metropolitan Life Insurance Company, New York.
2. *A Second Formula for Partial Sums of Hyper-geometric Series Having the Unit as Fourth Moment* Hermann von Schelling, Naval Medical Research Laboratory, New London, Connecticut.
3. *A Coverage Distribution.* Herbert Solomon, Office of Naval Research, Washington, D C
4. *The Problem of the Greater Mean.* R. R. Bahadur and Herbert Robbins, University of North Carolina

5. *Some Extensions of Bayes' Theorem.* F. C. Leone, Case Institute of Technology
6. *On Optimum Selections from Multinormal Populations.* Z. W. Birnbaum and D. G. Chapman, University of Washington.

On Wednesday morning, December 28, at 10:00 A. M. a session on *Cybernetics* was held jointly with the American Statistical Association and the American Mathematical Society The following papers were given:

1. *Technique of Multiple Prediction.* Norbert Wiener, Massachusetts Institute of Technology
2. *Stochastic Problems in Neurophysiology* Walter Pitts, Massachusetts Institute of Technology.
3. *Information Theory.* Claude Shannon, Bell Telephone Laboratories

with discussion by Professor J. L. Doob, University of Illinois, Professor Mark Kac, Cornell University, and Professor L. J. Savage, University of Chicago. Professor Jerzy Neyman, University of California was Chairman of the session.

The session on *Review of Statistical Methodology* was held jointly with the American Statistical Association at 2:00 P. M., Wednesday, December 28, with Professor W. A. Wallis, University of Chicago, as chairman. The two papers presented were: *Review of Statistical Methodology in Agriculture and Related Fields*, by Professor W. T. Federer, Cornell University and *Recent Developments in Statistical Methodology in Social Science*, by Professor Frederick Mosteller, Harvard University; discussion followed by Professor L. J. Savage of the University of Chicago.

The second session of contributed papers was held jointly with the American Statistical Association and the Econometric Society on Thursday, December 29, at 10:00 A. M., with Professor H. T. Davis of Northwestern University presiding. The following papers were presented:

1. *Simple Regression Analysis with Autocorrelated Disturbances.* Howard Jones, Illinois Bell Telephone Company.
2. *Application of Sequential Sampling Method to Check the Accuracy of a Perpetual Inventory Record* Joseph Jeming, New York City.
3 *A Test of Klein's Model III for Changes of Structure* Andrew Marshall, Rand Corporation.
4. *Application of the Theory of Extreme Values to Economic Problems* S. B. Littauer, Columbia University and E J. Gumbel, New York City.
5. *Bias Due to the Omission of Independent Variables in Ordinary Multiple Regression Analysis.* T. A. Bancroft, Iowa State College
6. *Estimating Parameters of Pearson Type III Populations from Truncated Samples* A C. Cohen, Jr , University of Georgia.
7. *The Circular Normal Distribution.* E. J. Gumbel, New York City.

The third session of contributed papers was held at 2:00 P. M. on Thursday, December 29, with Professor L. C. Aroian of Hunter College as Chairman. The following papers were presented in person or by title as indicated:

1 *Treatment of Attenuation Problems by Random Sampling.* H. Kahn and T. Harris, The Rand Corporation
2 *On the Existence of Nearly Locally Best Unbiased Estimates.* Herman Rubin, Stanford University.

3. *The Experimental Evaluation of Multiple Definite Integrals.* George Tyler, Naval Electronics Laboratory, San Diego, California.
4. *Tests of Fit of a Cumulative Distribution Function Over Partial Range of Sample Data.* Bradford Kimball, New York State Department of Public Service, New York City.
5. *Large Sample Tests for Comparing Percentage Points of Two Arbitrary Continuous Populations.* A. W. Marshall and John Walsh, The Rand Corporation.
6. *On the Distribution of Wald's Classification Statistics.* Harman L. Harter, Michigan State College.
7. *Analysis of Extreme Values.* W. J. Dixon, University of Oregon.
8. *A Note on the Variance of Truncated Normal Distributions.* (By title) A. C. Cohen, Jr., University of Georgia.
9. *Some Comments on the Efficiency of Significance Tests.* (By title) John Walsh, The Rand Corporation.
10. *Some Estimates and Tests Based on the Smallest Values in a Sample.* (By title) John Walsh, The Rand Corporation.

The subject of the next session, 4:00 P. M. Thursday, December 29, was the *Review of Stochastic Processes from the Point of View of Mathematical Statistics.* This session was held jointly with the American Statistical Association, Professor C. C. Craig of the University of Michigan presiding. Two papers were given, one by Professor A. B. Mann of the National Bureau of Standards, Ohio State University and the University of California; and the second by Professor John Tukey, Princeton University.

On Friday, December 30, at 9:00 A. M. a session on *Statistical Methods in Astronomy* was held jointly with the American Statistical Association and Section D of the American Association for the Advancement of Socience. Professor Walter Bartky of the University of Chicago, Chairman of the session, opened the meeting with introductory remarks on *Astronomical Problems Requiring Statistical Methods.* The following papers were presented:

1. *The Nearby Stars.* Peter Van De Kamp, Swarthmore College.
2. *Corrections to Observed Frequency Distributions.* Bart J. Bok and J. K. De Jonge, Harvard University.
3. *The Problem of Selective Identifiability of Binaries.* Elizabeth Scott, University of California.
4. *Multivariate Periodogram Analysis and Detection of Variable Stars.* Harold Hotelling, University of North Carolina.

These papers were discussed by Professor Jerzy Neyman, University of California.

The session on *Discriminant Functions in Education* was held jointly with the American Statistical Association, the American Psychological Association and the Psychometric Society. Professor T. W. Anderson of Columbia University gave an invited address on *Classification by Multivariate Measures*, followed by discussion by Professors J. C. Flanagan of the University of Pittsburgh and John Carroll of Harvard University. Professor Robert Thorndike of Columbia University presided.

The final session of the meeting was devoted to *Computation* and was held jointly with the American Statistical Association and the Association for Computing Machinery. Professor Harold Hotelling of the University of North Carolina serving as Chairman. The following papers were given:

1. *Idiosyncrasies of Automatically-sequenced Digital Computing Machines.* Ida Rhodes, National Bureau of Standards.
2. *Problem Solving on Large-Scale Automatic Calculating Machines.* W. D. Woo, Harvard University.
3. *A Statistical Application of the UNIVAC.* John Mauchly, Eckert-Mauchly Computer Corporation.

These papers were discussed by James McPherson, Bureau of the Census and Emil Schell, Office of the Air Comptroller.

Meetings of the Council were held on Tuesday, December 27, at 12:00 Noon, Professor Jerzy Neyman presiding and again on Thursday, December 29, at 12:00 Noon, Professor J. L. Doob presiding. The Business Meeting was held on Wednesday, December 28, Professor Jerzy Neyman presiding. The report of this meeting is given elsewhere in this issue.

S. B. LITTAUER,
Associate Secretary

MINUTES OF THE ANNUAL MEMBERSHIP MEETING, NEW YORK, DECEMBER 28, 1949

The meeting was called to order at 4:30 P.M. by President Jerzy Neyman. The annual reports of the President, Editor, and Secretary-Treasurer were read. They are printed elsewhere in this issue.

It was moved by Harold Hotelling that the front cover of the *Annals* in the future shall bear the additional notation that it was edited during the years 1938–1949 by S. S. Wilks. Motion was seconded and carried unanimously.

The tellers reported the election of the following officers:

President-Elect	P. S. Dwyer
Members of the Council for 1950–1952	David Blackwell
	W. G. Madow
	Frederick Mosteller
	L. J. Savage

Meeting was adjourned at 5:15 P.M.

CARL H. FISCHER
Secretary

REPORT OF THE PRESIDENT OF THE INSTITUTE FOR 1949.

I wish to begin my Report by welcoming the newly elected Fellows, Doctors Z. W. Birnbaum, D. J. Finney, H. O. Hartley, Wassily Hoeffding, Michel Loève, Edward Paulson and S. N. Roy. In addition, a hearty welcome is due to Dr. G. W. Brown who was elected last year, but inadvertently omitted in the published list. The election to the fellowship is a mark of recognition on the part of the Institute. At the same time, I am sure the Institute has reason to be proud of having among its fellows such distinguished scholars as are now added to the list.

During the past year the intensity of the Institute's life grew markedly in many respects. In particular, a very considerable number of our members took part in various Committees. For the sake of brevity, the composition of all the Committees is given in a tabular form at the end of the Report. At this time I wish to express the indebtedness of the Institute to the Chairmen and to the Members of all the Committees.

Undoubtedly the most important function of the members of the Institute is research and the most important function of the Institute itself is the publication of the results of this research. In this respect the past year brought about a fundamental change: after a dozen years of hard and most fruitful work as Editor of the *Annals*, Professor S. S. Wilks resigned this year and the Council elected Professor T. W. Anderson as his successor. According to our present Constitution, the term of office of the Editor is three years.

About a decade ago I suggested and the Membership Meeting of the Institute approved that the cover of the *Annals* bear the name of its founder, Professor Harry Carver. Founded by Carver, the *Annals* were developed by Wilks and, now stand as the most important statistical journal in the world. Accordingly the Chair will welcome a motion to add Professor Wilks' name as a permanent feature of the cover of the *Annals of Mathematical Statistics*.

While being grateful to Wilks and regretting his withdrawal, we should extend a most hearty welcome to T. W. Anderson. Because of his scholarship, broad vision combined with broadmindedness and because of his energy, he is an excellent promise for the future of the *Annals*. It is a pleasure to express the gratitude of the Institute to Columbia University and, in particular, to Dr. Abraham Wald for providing the necessary facilities for the Editorial office of the *Annals*.

Prior to embarking on the election of the new Editor, the I.M.S. Council approved an important document prepared by a special Committee chaired by S. S. Wilks, formulating the editorial policy of the Institute.

Of the many fundamental parts of this document I wish to mention the following:

(i) "In establishing the editorial procedure, special care should be taken to avoid the danger of the Annals becoming a one-group journal rather than serving the Institute as a whole . . . the refusal to publish a paper on grounds of general policy (rather than because of some verifiable defects such as mistakes, triviality, lack of new material, etc.) shall be based on a unanimous agreement of the Editor and of all the Associate Editors."

The general idea behind these passages is, of course, that thus far, the *Annals* is the only journal published by the Institute and should provide facilities for all the different schools of thought. My understanding is that this includes the biostatistician Cochran and the econo-statistician Koopmans, the multivariate Hotelling and the tolerant Wilks, the quality-control-minded Shewhart and the dependently-limiting Loeve, the necessary- and sufficient-normal Feller and the minimax-gambler von Neumann, the relativistically-cybernetic Wiener and the general-sequential-decision-maker Wald. I should think that even our next

President, the stochastically-processed-Markovian Doob, is meant to have a chance to publish in the *Annals*, from time to time.

(ii) Another interesting point in the same document concerns the proposed approximate distribution of space in the Annals:
   (a) research papers on mathematical statistics proper—60 per cent;
   (b) research papers in borderline fields, including applications—20 per cent;
   (c) expository papers—15 per cent
   (d) news, notices, etc.—5 per cent.

Since in the past there was too little expository material, the Council instituted the so-called Special Inivited Papers, to be presented from time to time on selected subjects. The text of these papers, accompanied by the prepared discussion, will be printed in the *Annals*. The program of the present meeting includes our first Special Invited Paper, by Michel Loève. It is hoped that the Special Invited Papers will satisfy the need for expository material now felt by the membership of the Institute. I am sure the Program Committees will appreciate suggestions of the Members regarding the sections of the theory requiring expository presentation.

The financial aspect of the publication program of the Institute was a continued worry of the Council. As is well known, the *Annals* is overloaded with papers and the cost of printing is growing constantly. In order to ease the situation somewhat, our new Constitution was amended to include the provision that the Universities and other institutions could become Institutional Members. There is already some additional income from this source and, if all the members of the Institute are energetic in urging their Departments to become Institutional Members, this income may be quite substantial.

It is conceivable that some potential sources of funds exist, not directly available for the *Annals*, which may be used for starting a new statistical journal. In order to investigate this possibility a special committee was appointed under the chairmanship of Professor Scheffé. This Committee did an excellent job in trying to find a solution of the tremendously difficult problem and there is now a reasonable hope that, in the not very distant future, our publication facilities will be increased.

Another deep change in the structure of the Institute occurred this year. Here I have in mind the resignation of Dr. Paul S. Dwyer, our long and hard working Secretary, and the taking over by Dr. Carl Fischer. Dr. Dwyer's resignation was announced last year at the meeting at Cleveland and we expressed to him our hearty thanks for his untiring work for the Institute. I wish to repeat these thanks now and to accompany them by the hearty congratulations on the excellent program he prepared for this meeting in his new capacity as the Chairman of the National Program Committee

Until recently, there was a certain disequilibrium in the location of the meetings of the Institute. Practically all of the meetings were held in the East and the West Coast members could attend them only as a matter of exceptional luck. Later, regional meetings were organized, and this year we have functioning three

Regional Program Committees, one for the East, one for the West Coast and one for the Middle West. In addition, we have Program Committees for the two National Meetings of the Institute. In parallel with the redistribution of meetings, there was an increase in their number. This process was accompanied by the very efficient help on the part of the governmental organizations, of the Office of Naval Research, the Air Force, and the Army, for the members of the Institute to attend the meetings even if they are held at a considerable distance. As a combined result of these developments it now may seem that there are too many meetings. Undoubtedly, the number and the location of future meetings of the Institute will be seriously discussed and adjusted to the existing needs.

Naturally, the help of the Governmental institutions was not limited to help in travel. A considerable number of research projects in statistics are now in progress in many institutions with excellent results for science, for the younger people who are given the chance to make their first independent research work without undue worry about food and shelter and, thus, for the country as a whole. The first organization to support fundamental research in general, and in statistics, in particular, seems to be the office of Naval Research. Its broadmindedness and understanding of the spirit of research have established a very high standard which is also sustained by other institutions. If permitted to function as they do now, these institutions will mark an epoch in the development of scholarly work in this country.

The following persons have accepted the appointment to the Nominating Committee for the next year

Henry Scheffé—*Chairman*
Albert W. Bowker
Paul G. Hoel
Leonid Hurwicz
Herbert E. Robbins
David F. Votaw, Jr.

Composition of the Committees of the Institute in 1949

1. *Program Committees* (P.C.)

(i) Eastern P.C. for the April 1949 meeting in New York
Churchill Eisenhart, *Chairman*
W. G. Cochran
C. F. Kossack
S. B. Littauer
F. Mosteller

(ii) West Coast P. C. for June meeting in Berkeley
M. A Girshick, *Chairman*
Z. W. Birnbaum
W. J. Dixon
J. L. Hodges, Jr.
P. G. Hoel
A. M. Mood

(iii) National P.C. for the Summer Meeting at Boulder, Colorado
W. Feller, *Chairman*

(iv) Mid West P.C.
C. C. Craig, *Chairman*

J. L. Doob
M. A. Girshick
C. C. Hurd
J. Wolfowitz
L. Hurwicz
W. G. Madow
K. May
L. J. Savage
D. R. Whitney

(v) National P.C. for the December meeting in New York
P. S. Dwyer, *Chairman*
J. Berkson
G. W. Brown
C. Eisenhart
Mark Kac
H. Rubin

(vi) Eastern P.C. for the Spring 1950 meeting in North Carolina.
H. Hotelling, *Chairman*
D. Blackwell
H. Geiringer
S. B. Littauer
D. F. Votaw, Jr.
S. S. Wilks

2. *Committee for Special Invited Papers*
J. W. Tukey, *Program Coordinator, Chairman ex officio*
C. C. Craig
P. S. Dwyer
C. Eisenhart
W. Feller
M. A. Girshick
H. Hotelling

3. *Committee on Editorial Policy (1948–1949)*
S. S. Wilks, *Chairman*
W. G. Cochran
W. Feller
M. A. Girshick
P. S. Olmstead
J. Neyman
W. A. Wallis
J. Wolfowitz

4. *Committee to Nominate Candidates for the Editor of the Annals*
Harry C. Carver, *Chairman*
David Blackwell
S. Lee Crump
Erich L. Lehmann
Howard Levene
Frederick Mosteller
Herbert E. Robbins

5. *Committee on Tabulation*
C. Eisenhart, *Chairman*
C. I. Bliss
F. W. Dresch
H. H. Germond
H. O. Hartley
C. C. Hurd
A. N. Lowan
W. G. Madow
H. G. Romig
L. E. Simon

6. *Committee on Directory*
John W. Tukey, *Chairman*
Churchill Eisenhart

7. *Committee to Revive the Statistical Research Memoirs*
Henry Scheffé, *Chairman*
T. W. Anderson
Walter Bartky
C. C Hurd
George Kuznets

8. *Rietz Lectures Committee*

   The Chairmanship of this Committee was accepted by Abraham Wald, the first Rietz Lecturer, who undertook to make further appointments. These are:

   C. C. Craig
   W. Feller
9. *Committee to Encourage Membership outside of the United States*
   T. W. Anderson, *Chairman*
   C. C. Hurd
   M. Loève
   J Marschak
10. *Committee on Statisticians in the Government Service*
    W. E. Deming, *Chairman*
    C. Eisenhart
11. *Representative of the I.M.S. to the American Association for the Advancement of Science*
    Harold Hotelling
12. *Representative of the I.M.S. to the National Research Council, Division of Physical Sciences*
    Walter Bartky (1948–1950)
13. *Representative of the I.M.S. to the Mathematical Policy Committee*
    S. S. Wilks
14. *Representative of the I.M.S. to the Joint Committee for Development of Statistical Applications in Engineering and Manufacturing*
    Benjamin Epstein
15. *Representatives to the Inter-Society Cooperation on Mathematical Training of Social Scientists*
    T. W Anderson
    J. L. Doob
    S. S. Wilks
16. *Committee to Determine the Duties and Responsibilities of the Program Committees*
    Harold Hotelling, *Chairman*
    M. A. Girshick
    S. B. Littauer

J. NEYMAN
President

December 31, 1949

REPORT OF THE SECRETARY-TREASURER OF THE INSTITUTE FOR 1949

At the beginning of 1949 the Institute had 1101 members and during the period covered by this report 153 new members (8 of whom begin their membership with 1950) joined the Institute and two members were re-instated. During 1949 the Institute lost 87 members of which 27 were [illegible]

suspension for non-payment of dues. Judging from the information available at this date, the Institute will have 1167 members as it starts 1950.

During 1949 the Constitution was amended to provide for a new class of membership: Institutional Membership. Although the campaign for institutional members started late in the year, by December 31 there were five universities on the rolls: California, Purdue, Illinois, Princeton and North Carolina. It is hoped that many more universities and corporations will enroll during 1950.

Meetings of the Institute held during 1949 included those at Columbia University on April 8–9, at the Berkeley campus of the University of California on June 16–18, at the University of Colorado on August 29–September 1, and at New York City on December 27–30. The Secretary wishes to call attention to the excellent work of the members who served as Assistant and Associate Secretaries at these meetings: Professor S. B. Littauer at New York, Professor J. L. Hodges, Jr., at California, Professor H. T. Guard at Colorado and Associate Secretary Professor Littauer who was responsible for the New York Meeting.

The following Fellows served as members of the Committee on Fellows: C. C. Craig, chairman, T. W. Anderson, M. A. Girshick, Harold Hotelling, Henry Scheffé, and F. F. Stephan.

The meeting scheduled for November 25–26 at the University of California at Los Angeles was cancelled by vote of the West Coast membership because of the proximity of the Boulder and Christmas Meetings.

At the Council meeting at Boulder, August 29, 1949, the following Associate Secretaries were elected:

Associate Secretary	*Section*
S. B. Littauer	Eastern
K J. Arnold	Central
J. L. Hodges, Jr.	Western

By a mail vote of the Council, conducted during October, 1949, T. W. Anderson was elected Editor for the period 1950–1952.

A summary of the financial status of the Institute is given below:

FINANCIAL STATEMENT

December 20, 1948 to December 31, 1949

A. RECEIPTS

Balance on Hand,* December 20, 1948	$ 7,121.01
Dues	7,826.35
Contributions	156 15
Life Memberships	392.50
Institutional Memberships	400.00
Subscriptions	4,779.07
Sale of Back Issues	3,314.41
Biometrika	793.50
Income from Investments	100.00
Miscellaneous	169.70
Total	$25,052 69

* In bank deposits and government bonds.

B. EXPENDITURES

Annals—Current		
Office of the Editor	$ 275.00	
Waverly Press	8,777 65	$ 9,052.65
Annals—Back Numbers		
Reprinted Vol. II #4; III #4, IV #3 & #4; V #1, VI #1, 2, 3 & 4; XIII #1, 2, & 4.		$ 2,910.55
Mathematical Reviews and Inter-Society Committee		206.02
Office of the Secretary-Treasurer		
Printing, memoranda, etc. (Including some stamped envelopes)	$1,156.61	
Postage, supplies, express, telephone calls	275 00	
Clerical help	2,208.40	
Travelling expense	223,61	$ 3,863.02
Miscellaneous		$ 379.57
Biometrika		$ 657.30
Balance on Hand, *December 31, 1949		$ 7,982.08
Total		$25,052 69

C. SUMMARY OF RECEIPTS AND EXPENDITURES

Balance on Hand, *December 20, 1948	$ 7,121.01
Receipts during 1949	17,931.68
Expenditures during 1949	17,070.61
Balance on Hand, *December 31, 1949	$ 7,982.08

D. LIFE MEMBERSHIP FUNDS

It has been the practice to set up an amount equal to all life membership payments as a liability and to hold all these funds in reserve until the death of the member—after which his payment is released to the general fund. There were three new life membership payments in 1949.

	December 20, 1948	*December 31, 1949*
Number of Life Members	29	32
Total Reserve Held	$2,280.00	$2,672.50

E. BACK ISSUES FUND

It has been our policy, since January 1, 1948, to use income from the sale of back issues to finance the additional reprinting of back issues.

Previous balance in back issues fund	$ 749.77
Income from the sale of back issues during 1949	3,314 41
Expense for reprinting back issues in 1949	2,910.55
Balance, December 31, 1949	$1,153.63

F. BALANCE SHEET, DECEMBER 31, 1949

ASSETS

	December 31, 1949	*Increase since December 20, 1948*
Cash	$ 3,094.08	$ 861.07
U S Government G Bonds	3,000.00	—
U.S. Government F Bonds (Purchase price)	1,888.00	—
Current Accounts Receivable	545.78	254.56
Estimated Value (Cost of Back Annals**)	16,459.22	3,673 61
	$24,987.08	$4,789.24

* In bank deposits and government bonds

** Cost of *Annals* calculated at 67 cents per copy

LIABILITIES		
Reserve for Life Memberships	$ 2,672.50	$ 392 50
Reserve for Reprinting Back Issues	1,153.63	403.86
Surplus	21,160.95	3,992.88
	$24,987.08	$4,789 24

G. SUMMARY

The surplus of the Institute has increased during the year of 1949 by $3,992.88. While this indicates a favorable condition, it should be noted that roughly 92% of this gain is represented by an increase in the inventory of back issues of the Annals. This asset is definitely of the non-liquid sort and thus the major portion of our gain is of little assistance in meeting our current need for more publication space in the *Annals*.

It should be noted that the year-end statements have always included a substantial amount in prepaid dues and subscriptions on the asset side without a corresponding liability. The figure for December 20, 1948 is $4,060.50 and for December 31, 1949 is $4,682 37. Thus it will be seen that we are virtually running on a hand-to-mouth basis. It is hoped that an increase in the number of individual and institutional memberships during 1950 will bring us into a more favorable situation

Beginning with January 1, 1950 we plan to revise the bookkeeping system which is no longer adequate for an organization of our present size. In the future, these reports will be made on an accural basis rather than a cash basis and thus will present the data pertaining to each year on a more realistic basis

We are now in a position to supply all issues beginning with Volume 1. Five or six of the back issues are in short supply, but we expect to be able to reprint these when our supplies become exhausted, using receipts from the sale of back issues to pay for the reprinting.

CARL H. FISCHER
December 31, 1949 *Secretary-Treasurer*

REPORT OF THE EDITOR OF THE ANNALS FOR 1949

The 1949 volume of the *Annals* exceeded, by a few pages, the 600 pages budgeted for it at the beginning of the year. A total of 65 papers were published, as well as the usual reports, abstracts, and items of news and notices. The 1949 volume was Volume 20 of the *Annals*, and it seemed fitting to publish a cumulative index of papers for the first twenty volumes of the *Annals*. Such an index, containing both author and subject indexes, has been published as a separate 31-page pamphlet and is being distributed with the December 1949 issue of the *Annals*.

The rate of submission of manuscripts continues to increase. By the end of 1949 enough manuscripts to fill two issues of the *Annals* had been accepted for publication. At the same time approximately forty manuscripts were at various stages of refereeing and revision. This means that authors submitting manuscripts at the beginning of 1950 can hardly expect to see their papers in print in less than a year. The rate at which the average gap between submission of manuscripts and their appearance in print has, for the last two years, increased about two issues (six months) per year. There is no reason to predict that this rate will change for at least another year or two. Thus, it is highly desirable that every effort be made to expand the publication program of the Institute during 1950.

The most immediate possibility would be to expand the *Annals* by at least 100 pages if the budget will permit. In the meantime, it is hoped that the Institute committee to study the feasibility of reviving the *Statistical Research Memoirs* will be able to work out a practical plan for further increasing the publication facilities of the Institute.

The manuscripts being submitted continue to cover a wide range of topics in probability and statistics. There is still a scarcity of good review and expository articles being submitted, but with the institution of special invited addresses so widely discussed at the Cleveland meeting of the Institute in December, 1948, we can expect to receive more review and expository articles in the future.

The Editor takes this opportunity to acknowledge, on behalf of the Editorial Committee, the refereeing assistance which has been generously given during the year by the following persons: A. C. Aitken, E. W. Barankin, Z. W. Birnbaum, R. C Bose, A. H. Bowker, G. W. Brown, K. L. Chung, W. J. Dixon, A. Dvoretzsky, Hilda Geiringer, L. A. Goodman. T. N. E. Greville, F. E. Grubbs, John Gurland, M. H. Hansen, T. E. Harris, H. O. Hartley, E. L. Kaplan, B. F. Kimball, T. Koopmans, Julius Lieblein, H. Levene, M. S. MacPhail, P. J. McCarthy, R. B. Murphy, G. E. Noether, E. G. Olds, P. S. Olmstead, Richard Otter, E. Paulson, M. P. Peisakoff, E. J. G Pitman, Milton Sobel, D. F. Votaw, Max Woodbury, and J. L. Walsh.

Thanks are due to Mr. M. E. Freeman, Mr. L. A. Goodman and Mr. E. F. Whittlesey for preparation of manuscripts and to Mrs. Lily D. Smith for other editorial and office assistance in connection with the *Annals*.

Finally, on behalf of the Editorial Board, which has had the responsibility for editing the *Annals* since 1938, the Editor extends every good wish to the new Editor, T. W. Anderson, and the new Editorial Board, who will inherit nearly a full year of accepted manuscripts but will otherwise assume editorial responsibility for the *Annals* beginning with the 1950 volume.

S. S. Wilks
Editor.

December 21, 1949

THE IDENTIFICATION OF STRUCTURAL CHARACTERISTICS[1]

By T. C. Koopmans and O. Reiersøl

Cowles Commission for Research in Economics

1. Introduction.

1.1. *"Population" versus "structure."* In a fundamental paper (Fisher, [1]) R. A. Fisher distinguished as the first group of problems in mathematical statistics the "specification of the mathematical form of the population from which the data are regarded as a sample." It is the purpose of this article to suggest a reformulation of the specification problem, appropriate to many applications of statistical methods, and to point out the consequent emergence of a new group of problems, to be called identification problems.

In many fields the objective of the investigator's inquisitiveness is not just a "population" in the sense of a distribution of observable variables, but a physical structure projected behind this distribution, by which the latter is thought to be generated. The word "physical" is used merely to convey that the structure concept is based on the investigator's ideas as to the "explanation" or "formation" of the phenomena studied, briefly, on his theory of these phenomena, whether they are classified as physical in the literal sense, biological, psychological, sociological, economic or otherwise. Examples of such structures, drawn from the fields of economic fluctuations and of psychological factor analysis, are given in sections 3 and 4. More detailed discussions of these examples can be found in other publications by the present authors and by others [15], [19]. In this article, we are therefore not concerned with the merits of particular assumptions entering into the specifications considered. Our examples are used only as the basis for a generalizing formulation (Section 2) and a comparative discussion (Section 5) of the identification problem, i.e., the problem of drawing inferences from the probability distribution of the observed variables to the underlying structure. The belief is here expressed that this is a general and fundamental problem arising, in many fields of inquiry, as a concomitant of the scientific procedure that postulates the existence of a structure.

The general formulation of the identification problem in Section 2 is, therefore, held abstract. Some readers may prefer to give substance to the various concepts by reading Sections 3–4 alongside Section 2. In addition, we insert here a simple example showing the main features of the identification problem.

---

[1] To be included in Cowles Commission Papers, New Series, No. 39. The authors reported on this study in papers before the Berkeley meeting of the Institute of Mathematical Statistics in June 1948. We are indebted to Dr. G. Rasch of the University of Copenhagen and to Professor L. L. Thurstone of the University of Chicago for many fruitful discussions on the subject matter of this article, for which the responsibility lies exclusively with the authors.

1.2. *A simple example of the identification problem.* This example is concerned with the problem of estimating the parameters α, β, of a linear relationship

$$\eta_2 = \alpha + \beta\eta_1 \tag{1.1}$$

between two variables η_1 and η_2 both of which are observed only subject to errors of observation u_1 and u_2. Thus, observations are available only for the variables

$$y_i = \eta_i + u_i \quad \text{where} \quad E(u_i) = 0, \qquad i = 1, 2. \tag{1.2}$$

The question under what conditions a consistent estimate of β exists has repeatedly attracted attention. To discuss this question, we shall consider a model in which η_1 is independent of (u_1, u_2) and in which the joint distribution of u_1 and u_2 is normal.

If also the distribution of η_1 is normal, it is easy to see that β cannot be determined from a knowledge of the joint probability distribution of the observed variables y_1 and y_2.[2] In this case the joint distribution of y_1 and y_2 is also normal and the distribution is completely characterized by five parameters, $E(y_1)$, $E(y_2)$, var (y_1), var (y_2), and cov (y_1, y_2). The parameters β and var (η_1) may now be chosen in any way such that the second term in the right hand member of

$$\begin{bmatrix} \text{var}(y_1) & \text{cov}(y_1, y_2) \\ \text{cov}(y_1, y_2) & \text{var}(y_2) \end{bmatrix} = \begin{bmatrix} 1 & \beta \\ \beta & \beta^2 \end{bmatrix} \text{var}(\eta_1) + \begin{bmatrix} \text{var}(u_1) & \text{cov}(u_1, u_2) \\ \text{cov}(u_1, u_2) & \text{var}(u_2) \end{bmatrix}$$

is a positive definite matrix. It is clear that if the left hand member is nonsingular, this condition can be met for any arbitrary value of β combined with a sufficiently small value of var (η_1).

It can be shown that β is uniquely determined by the joint probability distribution of y_1 and y_2 if this distribution is not normal. We shall prove this in the case that certain semi-invariants exist.[3]

Let $\phi_{y_1y_2}(t_1, t_2)$ denote the characteristic function of the joint distribution of y_1 and y_2

$$\phi_{y_1y_2}(t_1, t_2) = E(e^{y_1it_1 + y_2it_2}), \tag{1.3}$$

and let

$$\psi_{y_1y_2}(t_1, t_2) = \log \phi_{y_1y_2}(t_1, t_2). \tag{1.4}$$

Similar notations will be used for the characteristic functions of other random variables, and the logarithms of these functions.

Since (u_1, u_2) and (η_1, η_2) are independent, we obtain

$$\psi_{y_1y_2}(t_1, t_2) = \psi_{\eta_1\eta_2}(t_1, t_2) + \psi_{u_1u_2}(t_1, t_2), \tag{1.5}$$

---

[2] See [13], middle of page 70.

[3] The following proof is analogous to that given by Geary [8] in the case when the u's are not supposed to be normally distributed, but independent.

and from equations (1.1) and (1.3) we obtain

$$\phi_{\eta_1\eta_2}(t_1, t_2) = E(e^{\eta_1 it_1 + (\alpha+\beta\eta_1)it_2}) = e^{\alpha it_2}\phi_{\eta_1}(t_1 + \beta t_2),$$

or

$$(1.6) \qquad \psi_{\eta_1\eta_2} = \alpha it_2 + \psi_{\eta_1}(t_1 + \beta t_2).$$

Combining (1.5) and (1.6), we have

$$(1.7) \qquad \psi_{y_1y_2}(t_1, t_2) = \alpha it_2 + \psi_{\eta_1}(t_1 + \beta t_2) + \psi_{u_1u_2}(t_1, t_2),$$

where $\psi_{u_1u_2}(t_1, t_2)$ is a polynomial of second degree, since the joint distribution u_1 and u_2 is normal. Let κ_{rs} be the semi-invariants of the distribution of (y_1, y_2) and let κ_r be the semi-invariants of the distribution of η_1. Comparing coefficients in equation (1.7), we obtain

$$(1.8) \qquad \kappa_{rs} = \beta^s\kappa_{r+s} \qquad (r + s \geq 3)$$

and from this equation again

$$(1.9) \qquad \kappa_{rs} = \beta\kappa_{r+1,s-1} \qquad (r + s \geq 3, s \geq 1).$$

If at least one κ_{rs} with $r + s \geq 3$, is finite and different from zero (which implies that the joint distribution of y_1 and y_2 is not normal), β may be determined from one such equation given the joint distribution function of y_1 and y_2.

1.3. *Remarks on the history of the identification problem.* The identification problem has been discussed, in various terminologies and formulations, by quantitative thinkers in several fields. It is interesting to note that most of the contributions have come from researchers whose main attention was directed to particular fields of application. For this reason, perhaps, its general formulation was not attempted until recently.

In economics, contributions of increasing explicitness and generality were made by Pigou [18], Henry Schultz [20], Frisch [3], [4], [5], [6], [7], Marschak [17]. The main contributions to the formalization and explicit mathematical analysis of the problem were made so far by Haavelmo [9], Koopmans and Rubin [15], Wald [24], and Hurwicz [10].

In his books on factor analysis [21], [22], Thurstone discusses in several places questions of identifiability. Previously the lack of identifiability in a certain factor analysis model had been demonstrated by numerical examples by G. H. Thomson [27]. Models used in the analysis of latent structure in attitude and opinion research by Lazarsfeld [16] give rise to similar identification problems. In biometrics, the "method of path coefficients" of Sewall Wright [25], is essentially a method where a structure is postulated behind the observable distribution, and the identifiability of that structure discussed. The identification problem is also met with in the theory of the design of experiments, particularly in the method of confounding (Fisher [2], Chapter 7, Yates [26]). When con-

founding is used, the identifiability of certain parameters (second order interactions, say) is sacrificed in order to gain certain advantages in the testing of hypotheses concerning (and in the estimation of) the parameters that remain identifiable (main effects and first order interactions, say).

2. General formulation of the identification problem.

2.1. *Latent variables, observed variables, and structure.* In each of the examples considered in this article, the distributional specification applies directly to certain non-observable or in any case non-observed variables, variously referred to as errors of observation (like u_1 and u_2 above), disturbances, "true" variables (like η_1 above), specific factors, etc We shall refer to these as *latent variables*, denoted by a vector u. In addition, certain *structural relationships*—like (1.1) and (1.2)—are specified which connect the latent variables with the *observed variables*, denoted by a vector y The specification is therefore concerned with the mathematical forms of both the distribution of the latent variables and the relationships connecting observed and latent variables

The term "mathematical form" carries a suggestion of parametric specification which obviously is not the only possible type. We shall therefore employ terms and concepts introduced by Hurwicz [10] which cover both parametric and nonparametric specifications. By a *structure* $S = (F, \phi)$ we understand a particular probability distribution function

$$F(u) \tag{2.1}$$

of the latent variables—thought of, if you wish, as given numerically to a desired degree of accuracy, either by a cumulative distribution surface or curve or table, or parametrically by numerical values of the parameters—combined with a particular structural relationship (or set of simultaneously valid relationships)

$$\phi(y, u) = 0 \tag{2.2}$$

between observed and latent variables—again given numerically by curves, surfaces or parameters—which permits unique determination of the observed variables y from the values of the latent variables u (except possibly for a set of u-values occurring with probability zero) The corresponding probability distribution

$$H(y \mid S) \tag{2.3}$$

of the apparent variables is therefore uniquely determined by the structure S. and is said to be *generated* by S.

2.2. *Specification of a model.* We shall use the term *model* to signify a set of structures We can thus say that the specification problem is concerned with specifying a model[4] $\mathfrak{S}$ which by hypothesis contains the structure S generating the distribution H of the observed variables.

[4] A set will be denoted by a German character corresponding to the Latin character denoting its representative element.

As a result of this reformulation of the specification problem, a new problem of inference arises, which logically precedes all problems of estimation or of testing hypotheses. It has already been deduced from the definition of structure that a given structure S generates one and only one probability distribution $H(y \mid S)$ of the apparent variables. However, statistical inference from any number of observations can relate only to characteristics of the distribution of the observed variables. The limit of statistical inference is an exact knowledge of this distribution function, a limit not attainable but approachable if very large samples can be taken. Anything not implied in this distribution is not a possible object of statistical inference.

2.3. *Identifiability of structural characteristics by a model.* It is therefore a question of great practical importance whether a statement converse to the one just made is valid: can the distribution H of apparent variables, generated by a given structure S contained in a model $\mathfrak{S}$, be generated by only one structure in that model? This is by no means implied in the definitions given, and it is not generally true. Whether or not it is true in a particular instance depends—as illustrated in our examples—always on the model $\mathfrak{S}$, and often on the given structure S besides If it is true, we shall say that the model $\mathfrak{S}$ *identifies* the given structure S, or that the structure S is *identifiable* by the model.[5]

If a structure S is not identifiable by a model $\mathfrak{S}$, some of its characteristics may still be uniquely determinable By a *structural parameter* $\theta(S)$ we understand a functional of the structure S (This definition applies, of course, equally to the case of non-parametric specification of the functions F, ϕ defining the structure.) We further define that two structures S and S^* are (observationally) *equivalent* if they generate the same distribution of observed variables,

$$H(y \mid S) = H(y \mid S^*) \qquad \text{for all } y \tag{2.4}$$

We then say that a model $\mathfrak{S}$ identifies a parameter $\theta(S)$ in a structure S_0, if that parameter has the same value in all structures S_0^*, contained in $\mathfrak{S}$ and equivalent to S_0. This definition can obviously be extended to characteristics $\chi(S)$ of a structure S, other than parameters, such as the functional form of a relationship represented by a component of the vector ϕ, etc.

2 4. *The identification problem.* It has now become clear that our reformulation of the specification problem has given rise to a new group of *identification* problems: to determine which of the parameters or other characteristics of a given structure are identifiable by (or "within") a given model.

It is perhaps premature to attempt assigning to identification problems a definite place in a classification of statistical problems such as was undertaken by Fisher. One might regard problems of identifiability as a necessary part of the specification problem We would consider such a classification acceptable, provided the temptation to specify models in such a way as to produce identifiability of relevant characteristics is resisted. Scientific honesty demands that

[5] The concept here designated briefly as "identifiability" has been called "unique identifiability" in another context (Koopmans and Rubin [15], also Hurwicz [10]) in contrast with "multiple" or "incomplete" identifiability.

the specification of a model be based on prior knowledge of the phenomenon studied and possibly on criteria of simplicity, but not on the desire for identifiability of characteristics in which the researcher happens to be interested.

Identification problems are not problems of *statistical* inference in a strict sense, since the study of identifiability proceeds from a hypothetical exact knowledge of the probability distribution of observed variables rather than from a finite sample of observations. However, it is clear that the study of identifiability is undertaken in order to explore the limitations of statistical inference.

2.5. *Identifiability is subject to statistical test.* Further interpenetration of the pre-statistical analysis of identifiability with problems of statistical inference proper arises from the fact, amply illustrated by our examples, that the identifiability of a structural characteristic $\chi(S)$ often depends not only on the model, but also on the given structure S. Thus, each structural characteristic χ divides the model $\mathfrak{S}$ exhaustively into two mutually exclusive subsets of structures

$$\mathfrak{S} = \mathfrak{S}_\chi + \mathfrak{S}_{\bar\chi} \tag{2.5}$$

(of which one may be empty), such that $\chi(S)$ is uniquely identifiable in S_0 by the model if S_0 belongs to $\mathfrak{S}_\chi$, and not uniquely identifiable if S_0 belongs to $\mathfrak{S}_{\bar\chi}$. We shall call $\chi(S)$ *uniformly identifiable* by $\mathfrak{S}$ if $\mathfrak{S}_\chi$ coincides with $\mathfrak{S}$.

The subdivision of $\mathfrak{S}$ into $\mathfrak{S}_\chi$ and $\mathfrak{S}_{\bar\chi}$ has an important property: If S_0 belongs to $\mathfrak{S}_\chi$, then all structures S_0^* equivalent to S_0 also belong to $\mathfrak{S}_\chi$, and a similar statement holds for $\mathfrak{S}_{\bar\chi}$. This property follows directly from the definition of identifiability of $\chi(S)$ given above. Its meaning is that the identifiability of $\chi(S)$ in S_0 depends only on the distribution of $H(y) = H(y \mid S_0)$ of observed variables generated by S_0. To the subdivision of the model corresponds an exhaustive subdivision

$$\mathfrak{H} = \mathfrak{H}_\chi + \mathfrak{H}_{\bar\chi} \tag{2.6}$$

of the set

$$\mathfrak{H} = \mathfrak{H}(\mathfrak{S}) \tag{2.7}$$

of all distribution functions $H(y \mid S)$ generated by the structures S of $\mathfrak{S}$, into the subset $\mathfrak{H}_\chi$ containing those distribution functions $H(y \mid S)$ generated by structures S in which $\chi(S)$ is uniquely identifiable, and the subset $\mathfrak{H}_{\bar\chi}$ containing functions $H(y \mid S)$ generated by structures for which the opposite is true.

Hence, whenever the identifiability of $\chi(S)$ cannot be decided in the same sense (affirmatively or negatively) for all structures S of $\mathfrak{S}$ as a result of either $\mathfrak{S}_\chi$ or $\mathfrak{S}_{\bar\chi}$ being empty, then the identifiability of the characteristic $\chi(S)$ of the structure S generating the observations is a property of the distribution $H(y \mid S)$ of the observations. This identifiability is equivalent to the hypothesis

$$H(y \mid S) \quad \text{belongs to} \quad \mathfrak{H}_\chi, \tag{2.8}$$

which is in principle[6] subject to statistical test under the maintained hypothesis

(2.9) $$H(y \mid S) \quad \text{belongs to} \quad \mathfrak{H}.$$

2.6. *Testing particular specifications.* Often the model is defined by one general specification supplemented with a number of particular specifications which are "detachable pieces" in the sense that they can be removed, added or replaced by alternatives to construct alternative models. We may define the *general specification* as a set $\mathfrak{S}$ of structures which is postulated to contain the model $\mathfrak{S}'$ in question as a subset. *Particular specifications* can then be defined as subsets $\mathfrak{S}_1, \mathfrak{S}_2, \cdots$ of $\mathfrak{S}$ of which the model $\mathfrak{S}'$ is the intersection

(2.10) $$\mathfrak{S}' \equiv \mathfrak{S} \cap \mathfrak{S}_1 \cap \mathfrak{S}_2 \cap \cdots .$$

An example is that of parametric specification of the "form" of the functions $\phi(y, u)$ defining the structural relationships and of the distribution function $F(u)$ of latent variables as the general specification, and specifications of the values of certain parameters of ϕ and F as particular specifications.

In such situations, it is an important question whether a given particular specification is—again in principle—subject to statistical test. Whenever the answer depends on the other particular specifications, we may ask further which minimum set of other particular specifications must (together with the general specification) be entered into the "maintained hypothesis" in order that that given particular specification be subject to statistical test. A formal answer to this question, facilitating specific answers in each concrete case, can be given as follows.

Let a model $\mathfrak{S}$ be narrowed down to an alternative model

(2.11) $$\mathfrak{S}' = \mathfrak{S} \cap \mathfrak{S}_1$$

by a particular specification $\mathfrak{S}_1$. This particular specification will be called *observationally restrictive* if the set $\mathfrak{H}(\mathfrak{S}')$ of all distribution functions $H(y \mid S')$ of observed variables generated by the structures S' of $\mathfrak{S}'$ is a proper subset of the set $\mathfrak{H}(\mathfrak{S}')$ of all distribution functions $H(y \mid S)$ generated by the structures S of $\mathfrak{S}$. A statistical test of the particular specification $\mathfrak{S}_1$ can then be constructed by choosing as the hypothesis subject to test

(2.12) $$H(y) \quad \text{belongs to} \quad \mathfrak{H}(\mathfrak{S}'),$$

and as the maintained hypothesis

(2.13) $$H(y) \quad \text{belongs to} \quad \mathfrak{H}(\mathfrak{S}).$$

The particular specification $\mathfrak{S}_1$ remains subject to test if the model $\mathfrak{S}$ is stripped of such other particular specifications which are not necessary for the observationally restrictive character of $\mathfrak{S}_1$, although of course the outcome of the test may become either less or more certain as a result.

[6] See sub-section 2.7 below.

A frequent case of an observationally restrictive specification is that where a parameter $\theta(S)$ already identifiable in almost all structures S of $\mathfrak{S}$, is restricted by $\mathfrak{S}_1$ to a prescribed value (or to a prescribed point set not containing all points of its domain for all S of $\mathfrak{S}$) In this case, the specification in question has been called *overidentifying*.

2 7. *Remarks on the testing of hypotheses.* In subsections 2.5 and 2.6 we have without further inquiry applied the expression "hypothesis in principle subject to test" to any hypothesis which narrows down the set $\mathfrak{H}$ of distribution functions H generated by structures of the model to a proper subset $\mathfrak{H}'$. It will be clear that, to make a test actually possible, $\mathfrak{H}'$ cannot be allowed to be everywhere dense in $\mathfrak{H}$. For instance, if $\mathfrak{H}$ is defined parametrically, a hypothesis restricting $\mathfrak{H}'$ to rational values of the parameters is clearly not subject to statistical test. Just what set-theoretical requirements on $\mathfrak{H}'$ are needed to make a test possible is a separate problem which we shall not attempt to discuss.

We have also in another sense oversimplified the problem of testing particular specifications. In practice this problem presents itself as the choice of one out of many possible combinations of several particular specifications, rather than a number of separate and unconnected choices between the rejection and the adoption of each particular specification under consideration. Present theory of choice between two alternatives does not meet this situation.

3. An econometric example.[7]

In econometric studies[8] economic fluctuations have been described by a system of difference equations in (observed) economic variables y, subject to two kinds of outside influences, emanating respectively from (observed) exogenous -i.e., non-economic—variables z, and from (latent) random disturbances u. Each of these equations is given a definite meaning in terms of economic behavior. There may for instance be equations explaining respectively consumption expenditure (from incomes of various groups, price changes, etc.), the supply of consumers' goods (from price margins between such goods and their raw materials and labor, productive capacity, etc.), investment expenditure, the supply of capital goods, etc The purpose of the identification discussion is to investigate whether, on the basis of given a priori knowledge as to the form of these equations, and in particular as to what variables occur in any designated equation, procedures of estimation or testing of hypotheses can be directed to the parameters of the equations of economic behavior themselves, rather than to the parameters of "secondary" equations dependent on (derivable from) two or more of the behavior equations.

In the case of linear systems of equations, a possible form for the general specification (the model $\mathfrak{S}$) is as follows.

$$(3.1) \quad \mathrm{B}_0 y'(t) + \mathrm{B}_1 y'(t-1) + \cdots + \mathrm{B}_{\tau_{\max}} y'(t-\tau_{\max}) + \Gamma z'(t) = u'(t)$$

[7] For an expository discussion of identification problems in econometric models see [14]

[8] See, for instance, J. Tinbergen [23] and L. R. Klein [12].

represents the structural relationships. Here $y'(t)$, $z'(t)$, $u'(t)$ are column vectors (the transposes of row vectors) of G, K and G elements, respectively, for each discrete time point or period $t = 1, 2, \cdots, T$, also $t = 0, -1, \cdots, 1 - \tau_{max}$, for $y'(t)$ $B_0, B_1, \cdots, B_{\tau_{max}}$ are square matrices of order G, and Γ is a matrix of G rows and K columns

(3.2) B_0 is non-singular.

(3.3) The observed values $z(t)$, $t = 1, \cdots, T$, are held constant in repeated samples, and the components of $z(t)$ are linearly independent.

(3.4) The components of $u(t)$ have a joint distribution function $F(u)$ (with zero means and finite variances) which is independent of t and of $z(t)$.

(3.5) $u(t)$ and $u(t')$ are independently distributed if $t \neq t'$.

Particular specifications $\mathfrak{S}_1, \mathfrak{S}_2, \cdots$, that have been most frequently employed indicate prescribed values (usually zero) of specified elements of the matrix

$$A \equiv [B_0 \ B_1 \cdots B_{\tau_{max}} \ \Gamma] \tag{3.6}$$

or of given linear functions of the elements of the g^{th} row $\alpha(g)$ of A, for each value $g = 1, \cdots, G$ of g. It can always be arranged that of the linear restrictions on any one row of A, at most one is non-homogeneous (normalization rule), the others homogeneous. The homogeneous restrictions state which variables enter into each equation, and possibly with which ratios between some of their coefficients.

It has been shown [15] that in the model $\mathfrak{S}$, a necessary and sufficient condition for the equivalence of two structures $S \equiv \{F(u), A\}$ and $S^* \equiv \{F^*(u^*), A^*\}$ is that they are connected by a linear transformation

$$A^* = \Upsilon A, \qquad u'^* = \Upsilon u', \tag{3.7}$$

with non-singular matrix Υ. By definition, the model

$$\mathfrak{S}' = \mathfrak{S} \cap \mathfrak{S}_1 \cap \mathfrak{S}_2 \cap \cdots \tag{3.8}$$

identifies a parameter α_{gk} if, whenever A and A* belong to equivalent structures S and S^*, respectively, of $\mathfrak{S}'$, we have

$$\alpha_{gk}^* = \alpha_{gk} . \tag{3.9}$$

In order to attain such identifiability by linear restrictions on the g^{th} row of A it is necessary that one non-homogeneous restriction (normalization rule) on the g^{th} row of A be specified in $\mathfrak{S}'$. Recalling that G represents the number of rows (and the rank) of A, it can be proved that it is further necessary for the simultaneous identifiability of all elements α_{gk}, $k = 1, \cdots, K$, in the g^{th} row $\alpha(g)$ of A, that at least $G - 1$ additional non-homogeneous restrictions be imposed on that row, say

$$\alpha(g)\Phi'(g) = 0, \qquad \rho\{\Phi'(g)\} \geqq G - 1, \tag{3.10}$$

where $\alpha(g) \equiv [\alpha_{g1} \cdots \alpha_{gK}]$, the $\Phi(g)$ are given matrices (often with elements 0 or 1 only), and $\rho(X)$ denotes the rank of X. These restrictions (3.10) are also sufficient (in addition to the normalization rule) if

$$(3.11) \qquad \rho\{A\Phi'(g)\} = G - 1.$$

The g^{th} row of the "rank criterion matrix" $A\Phi'(g)$ in (3.11) consists of zeros only, because of (3.10). Therefore, (3.11) requires the other rows of that matrix to be linearly independent.[9]

Thus, even if the model $\mathfrak{S}'$ includes, besides a normalization rule, the necessary condition (3.10) for the identifiability of the g^{th} behavior equation, such identifiability is still absent in certain structures, corresponding to a point set (generally of measure zero) in the space of the coefficients of the remaining equations, viz., the point set in which (3.11) is not satisfied. Whether or not A actually falls within this point set is, as was stated before in more general terms, a property of the joint distribution function $H(y \mid z)$ of the observations y, and is therefore subject to statistical test. In the present case, this is also seen from the fact that the rank of $A\Phi_g'$ is preserved by the transformation (3.7), and is therefore itself an identifiable parameter.

For certain scientific purposes explicit knowledge of A is unnecessary. One such purpose is "prediction without change in structure," i.e., prediction of a value of $y(t)$ for a future time t from a hypothetical value of $z(t)$ on the assumption that A and $F(u)$ have not changed between the observation period and the time point to which the prediction applies. Such prediction can be based on the knowledge of (a) the population regressions

$$(3.12) \qquad y'(t) = \Pi_1 y'(t-1) + \cdots + \Pi_{\tau_{\max}} y'(t-\tau_{\max}) + \Pi_z z'(t) + v'(t)$$

of the "jointly dependent" variables $y(t)$ on the "predetermined" variables $y(t-1), \cdots, y(t-\tau_{\max})$, $z(t)$ and of (b) the distribution function $K(v)$ of the population residuals

$$(3.13) \qquad v(t) = y(t) - E\{y(t) \mid y(t-1), \cdots, y(t-\tau_{\max}), z(t)\}$$

from these regressions. Of course, the matrices "Π" are functions of the structural parameters (3.6) through

$$(3.14) \qquad [-I\ \Pi] \equiv [-I\ \Pi_1 \cdots \Pi_{\tau_{\max}}\ \Pi_z] = -B_0^{-1}A$$

and $K(v)$ can be derived from $F(u)$ through the transformation

$$(3.15) \qquad v' = B_0^{-1}u'.$$

The important fact is that Π and $K(v)$, by their definitions, depend only on the distribution function $H(y \mid z)$ of the observations, and are therefore uniformly identifiable. This is also reflected in the fact that the right hand members of (3.14) and (3.15) are invariant for the transformation (3.7).

[9] In that case, overidentification of $\alpha(g)$ will result if the inequality sign in (3.10) holds.

However, the most relevant economic problems are those in which a change in A or $F(u)$ is actually or hypothetically present, and in which therefore the identifiability of the relevant parts or functions of A and of the characteristics of $F(u)$ requires separate inquiry.[10]

4. An example from factor analysis.[11] Factor analysis has been presented in different forms by different authors We shall here consider the multiple factor analysis of Thurstone only [21], [22].

The factor analysis methods were developed primarily for the purpose of analyzing intelligence tests, but they have also been used for other psychological problems and in other sciences.

Suppose that a person is given a battery of G tests. Let his score in test i be y_i. The fundamental assumption in factor analysis is that these scores can be explained in terms of a relatively small number of hypothetical primary factors. Let $z_1, z_2, \cdots, z_\rho$ denote the hypothetical scores of the person in the common factors, i e., those primary factors which are common to at least two tests in the battery. We assume that y_i is a homogeneous linear function of the scores z_k plus a unique part v_i, which may be thought of as consisting of an error term plus the contribution of a specific factor. The coefficients π_{ik} in the linear function just mentioned are called factor loadings. The factor loading π_{ik} expresses the relative importance of the common factor k in the answering of test i.

We shall introduce the row vectors $y = [y_i]$, $z = [z_i]$, $v = [v_i]$ and the matrix $\Pi = [\pi_{ik}]$. The covariance matrices of the sets of variables y, z, and v will be denoted by M_{yy}, M_{zz}, and Δ, respectively.

In contrast with the preceding example, the variables y are the only observed variables. The variables v and z are latent variables.

Our model will be given by the following specifications:

(4 1) $y' = \Pi z' + v'$.

(4.2) $E(z) = 0$ and $E(v) = 0$.

(4 3) The set of variables z is stochastically independent of the set of variables v.

---

[10] See Hurwicz [11].

[11] Proofs of the statements in this section will be found in a separate paper by one of the authors (Reiersøl [19]). It should be noted that the notation is different in the two papers In the separate paper the notation is close to that of Thurstone. In the present paper the notation has been chosen to correspond in some way to the notation in the econometric example A list of corresponding symbols in the present paper and in Thurstone's books follows.

Present paper	y_i	z_k	π_{ik}	G	ρ	M_{yy}	M_{zz}	Δ
Thurstone	s_j	x_m	a_{jm}	n	r	R_1	R_{pq}	$R_1 - R$

It should be noted that M_{yy}, M_{zz}, and Δ are covariance matrices of the original variables. while R_1, R_{pq}, and R are covariance matrices of standardized variables.

(4.4) Δ is diagonal and different from 0.

(4.5) The elements of z and v are jointly normally distributed.

(4.6) Each y_i is correlated with at least one of the other y's.

(4.7) The rank of Π equals the number ρ of its columns.

(4.8) M_{zz} is nonsingular.

(4.9) ρ is the smallest number of variables z which is compatible with the joint probability distribution of the observed variables y and specifications (4.1)-(4.8).

(4.10) Each column of Π contains at least ρ zeros (in unspecified places).

(4.11) A normalization rule fixing the units of the variables x and a rule fixing the order of the columns of Π.

Denote by Π_k the matrix consisting of all the rows of Π which have a zero in the k^{th} column. Let the number of rows in the matrix Π_k be p_k. Let Π_{ki} denote the submatrix of Π_k which we get when deleting the i^{th} row of Π_k. Using these notations we shall formulate the final specification of our model.

(4.12) The rank of each of the matrices Π_{ki} $(k = 1, 2, \cdots, \rho; i = 1, 2, \cdots, p_k)$ is $\rho - 1$.

Specification (4.1) represents the structural relationships

Specification (4.10) means that the experimenter thinks he can construct a sufficient number of tests where at least one of the common primary factors is absent.

We shall first consider a model $\mathfrak{S}$ containing Specifications (4.1)-(4.9) only. From (4.9) follows that ρ is uniformly identifiable.

Let $\rho_G = \frac{1}{2}(2G + 1 - \sqrt{8G + 1})$. If $\rho > \rho_G$, the matrix Δ is generally not identifiable If $\rho < \rho_G$, Δ generally is identifiable When $\rho = \rho_G$, the number of values of Δ, which correspond to a given covariance matrix M_{yy}, is usually finite, and may be equal to one or greater than one. The matrices Π and M_{zz} are never identifiable in the model $\mathfrak{S}$. If Δ is identifiable, the set of all structures $\{\Pi^*, M_{zz}^*, \Delta\}$ equivalent to the structure $\{\Pi, M_{zz}, \Delta\}$ is given by the set of all matrices

$$\Pi^* = \Pi\Psi \tag{4.13}$$

and

$$M_{zz}^* = \Psi^{-1}M_{zz}(\Psi')^{-1}, \tag{4.14}$$

where Ψ is any square, ρ-rowed and nonsingular matrix.

In the following we shall confine our discussion to the case $\rho < \rho_G$, and to structures in which the matrix M_{yy} is such that Δ is identifiable in $\mathfrak{S}$.

We shall now consider the model $\mathfrak{S}'$ defined by Specifications (4.1)–(4.11). In this model a necessary and sufficient condition for the identifiability of Π is that any square ρ-rowed minor of Π which is of rank $\rho - 1$ is contained in one of the matrices Π_k. This condition excludes the possibility that all elements belonging to the intersection of $\rho - 1$ rows and two columns of Π are all equal to zero. In order to be able to use this result, the experimenter would have to be able to construct tests where one, but not more than one, common factor would be absent. Therefore the result is not particularly useful. In order not to exclude the case where two common factors occur in more than $\rho - 2$ tests, we have introduced Specification (4.12).

We shall finally consider the model $\mathfrak{S}''$ defined by Specifications (4.1)–(4.12). Assuming M_{yy} known, we can determine some value Π^* of Π which satisfies Specifications (4.1)–(4.9). Since, by assumption, Δ is identifiable in $\mathfrak{S}$, Π^* must be of the form $\Pi\Psi$, where Π is the true factor loadings matrix and Ψ is non-singular. Let Π_k^* be a submatrix of Π^* containing all the columns of Π^* and satisfying the following conditions

(4.15) The rank of Π_k^* is $\rho - 1$.

(4.16) The addition to Π_k^* of a row contained in Π^* but not in Π_k^* increases the rank to ρ.

(4.17) Each submatrix of Π_k^* obtained by deleting one row of Π_k^* has rank $\rho - 1$.

A necessary and sufficient condition for the identifiability of Π in the complete model $\mathfrak{S}''$ is that there exist exactly ρ submatrices Π_k^* of Π^* which satisfy conditions (4.15)–(4.17), and that the ρ vectors q_k, satisfying the equations $\Pi_k^* q_k = 0$ when $k = 1, 2, \cdots, \rho$, are linearly independent.

It should be noted that Specifications (4.10) and (4.12) are observationally restrictive, i.e., they are in principle subject to statistical test.

5. A comparative discussion of the examples given. Some comparative remarks on the three examples given in sections 1.2, 3 and 4 may illustrate our general discussion of the identification problem, given in section 2.

In each of the three examples considered, the model contains a general specification prescribing a parametric form of the structural relationships (2.2). Further particular specifications therefore take the form of parameter specifications in the function $\phi(y, u)$ in (2.2) and possibly in the distribution function (2.1) of latent variables. A comparison of the three examples shows a striking formal similarity of the identification problems to which they give rise. This similarity justifies our speaking of identification problems as a separate group of problems preparatory to statistical inference, of quite widespread occurrence. The same definitions of structure, model, parameter, identifiability are applicable and useful in each example In all three cases, parameters occur, the identifiability of which depends on other identifiable structural characteristics (the normality of a distribution function in one case, the ranks of parameter matrices in the other two cases).

Our remaining remarks will be drawn from the econometric and factor analysis examples only, partly because these illustrate the identification problem in greater elaboration, partly because the closer similarity of these examples permits us to notice interesting differences in greater detail.

Let us consider the particular case of the econometric example when there are no time lags between the y's in the structural relationships, i.e., when $\tau_{..} = 0$. In this case the *reduced form* (3.12) in the econometric example is of the same form as equation (4.1), which defines the structural relationships in the factor analysis example. The notation in the factor analysis example has been chosen with this similarity in mind. However, it should be emphasized that, while the variables y are observed in both examples and the variables u are latent in both examples, the variables z are observed in the econometric example and latent in the factor analysis example, and even the number of variables z is an unknown parameter ρ in the latter example. For this reason, the discussion of the identifiability of Δ in factor analysis has no counterpart in the econometric model. Furthermore, the identifiability of the matrix Π, which is automatic and uniform in the econometric model $\mathfrak{S}_e$, say, requires detailed specifications in the factor analysis model $\mathfrak{S}_f$, say, including the diagonality of Δ and prescriptions about the number of zero elements in each column.

The observability of z in the econometric case is exploited to postulate, behind the reduced form (3.12), a structure $\{F(u), A\}$ to be identified (where possible) from further specifications based on economic theory. Here we meet with another analogy, with differences, between the identification problem of A in $\mathfrak{S}_e$ and that of Π (given Δ) in $\mathfrak{S}_f$. In the latter problem, the set of matrices Π^*, belonging to a set of equivalent structures, is given by equation (4.13). This equation is analogous to the first of the equations (3.7) in the econometric case, with Π in $\mathfrak{S}_f$ now corresponding to A' in $\mathfrak{S}_e$.

If we were to specify zeros in assigned places in the factor loadings matrix Π, and to introduce a normalization rule for each column of Π, the results quoted in the econometric example would immediately be applicable to the factor analysis case. A necessary condition for the identifiability of Π, given that of Δ, would be that the number of specified zeros in each column of Π be at least $\rho - 1$. Necessary and sufficient for identifiability would be that the matrix consisting of all rows of Π which have specified zeros in the k^{th} column, be of the rank $\rho - 1$, for each value of k.

However, instead of specifying that given elements of Π be equal to zero, Thurstone assumes that we know that there is a certain minimum number of zeros in each column, but that we do not know which particular elements are zero. The specification of a certain number of zeros in undesignated places obviously represents a weaker assumption than the specification of the same number of zeros in designated places. It is therefore not surprising that the specification of $\rho - 1$ zeros in undesignated places in each column is never sufficient for identifiability of Π. Thus, in the model $\mathfrak{S}_f$, we have introduced the stronger specification (4.10). We have seen that even this specification is too weak to be practically

useful, and have introduced the additional Specification (4.12), which makes the factor analysis model still more different from the econometric model.

Continuing the analogy in which Λ' in $\mathfrak{S}_c$ corresponds to Π in $\mathfrak{S}_f$, we note an important feature common to both examples, and present in other situations as well Even if specifications sufficient, in number and variety of "points of application," for the identifiability of all structural parameters cannot be derived from a priori considerations, it remains possible to construct uniformly identifiable functions of these parameters, knowledge of which constitutes scientific information of more limited usefulness.

In the econometric example we have already seen that for certain purposes a knowledge of the uniformly identifiable matrix Π of the reduced form is sufficient, while for other purposes we need to know the matrix A. As a further illustration, suppose that we want to test for persistence of the structure by comparing the equation systems which we estimate from data for two different periods. Disregarding errors of estimation (which are not our present topic), if A is the same in both cases, Π will also be the same in both cases. It is therefore possible to arrive at a rejection of the persistence hypothesis by determining Π in both cases. Suppose next that one row (or several rows) of A are different in the two periods, while the other rows of A are identical in the two cases. If B_0 changes from one period to the other, we may expect each element of Π to change. If we can determine A for each period, the equality (as between periods) of some of the rows of A will indicate precisely the extent of validity of the persistence hypothesis. If we cannot determine A but only Π in each case, this verification will be lost.

Similarly, it may in factor analysis be sufficient for some purposes to consider what we may call the reduced form of Π. Let Π_I be the upper square part of Π which we shall assume to be nonsingular. The matrix $\Lambda = \Pi\,\Pi_I^{-1}$ will be called the reduced form of Π. It will be of the form $\begin{bmatrix} I \\ \Lambda_{II} \end{bmatrix}$. Λ is always identifiable when Δ is identifiable.

Suppose now that the same battery of tests is given to two different populations. Suppose that some of the factor loadings are different in the two populations, while other factor loadings are the same. If at least one of the different factor loadings occurs in the matrix Π_I, then each element of Λ_{II} may be expected to change, and the partial identity of the two structures cannot be discovered if we determine Λ only and not Π. On the other hand, if Π is the same in both cases, also Λ will be the same in both populations.

Let us next consider two different batteries given to the same population. We shall suppose that the two batteries have some tests in common. For each test which is common to the two batteries we ought to find the same factor loadings in both batteries. In other words, the matrices Π in the two cases ought to be partly identical. On the other hand, if Π_I contains rows corresponding to tests which are not common to the two batteries, the matrices Λ_{II} will be entirely different in the two cases. Therefore, again, identification of Π will be necessary to verify the equality of the factor loadings of tests common to both batteries.

A final remark relates to observationally restrictive specifications. Particularly where the model is to a large degree speculative, empirical confirmation of the validity or usefulness of the model is obtained only to the extent that observationally restrictive specifications are upheld by the data. Thus, Thurstone emphasizes that the number of factors ρ should be well below the value ρ_0 found above to be necessary in general for the identifiability of Δ, before a factor analysis can be regarded as successful (Thurstone [22], p. 294).

In econometric work, greater reliance is sometimes placed on a priori specification of the form of a behavior equation, particularly the variables occurring in it. If the linear restrictions on an equation in a linear system are just sufficient for its identifiability, estimation of the parameters of that equation is possible, but none of the identifying restrictions are themselves subject to test. Again, dependence on a priori information is diminished (but not eliminated) to the extent that a greater number of overidentifying restrictions are imposed and are upheld by the data.

REFERENCES

[1] R. A. FISHER, "On the mathematical foundations of theoretical statistics," *Phil. Trans. Roy Soc.*, London, Ser. A, Vol. 222 (1922), p. 309.

[2] R. A. FISHER, *The Design of Experiments*, Oliver and Boyd, Edinburgh and London, 1935

[3] R. FRISCH, "Correlation and scatter in statistical variables," *Nordic Stat. Jour.*, Vol. 1 (1928), p. 36.

[4] R. FRISCH AND B. D MUDGETT, "Statistical correlation and the theory of cluster types," *Jour. Am. Stat. Assn.*, Vol. 26 (1931), p. 375.

[5] R. FRISCH, "Pitfalls in the statistical construction of demand and supply curves," *Veröffentlichungen der Frankfurter Ges. für Konjunkturforschung*, Neue Folge, Heft 5, Leipzig 1935.

[6] R FRISCH, *Statistical Confluence Analysis by Means of Complete Regression Systems*, Publ. No. 5, Universitetets Økonomiske Institutt, Oslo, 1934

[7] R. FRISCH, *Statistical Versus Theoretical Relations in Economic Macrodynamics*, mimeographed document for League of Nations conference, 1938.

[8] R. C. GEARY, "Inherent relations between random variables," *Proc. Roy Irish Acad.*, Vol. 47 (1942), Sect A, No. 6.

[9] T. HAAVELMO, "The probability approach in econometrics," *Econometrica*, Vol. 12 (1944), Suppl.

[10] L HURWICZ, "Generalization of the concept of identification," *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10, New York, John Wiley and Sons, 1950.

[11] L. HURWICZ, "Prediction and least-squares," ibid.

[12] L. R. KLEIN, "Economic fluctuations in the United States, 1921–1941," *Cowles Commission Monograph 11*, New York, John Wiley and Sons, 1950.

[13] T. C. KOOPMANS, *Linear Regression Analysis of Economic Time Series*, Netherlands Economic Institute, Publ No. 20, Haarlem, 1937.

[14] T. C. KOOPMANS, "Identification problems in economic model construction," *Econometrica*, Vol 17 (1949), p 125

[15] T C. KOOPMANS, H RUBIN, AND R B. LEIPNIK, "Measuring the equation systems of dynamic economics," *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10, New York, John Wiley and Sons, 1950.

[16] P F LAZARSFELD, "The logical and mathematical foundation of latent structure analysis" "The interpretation of some latent structures." *Measurement and prediction*, Vol 4, Studies in Psychology of World War II, 1950

[17] J MARSCHAK, "Economic interdependence and statistical analysis," *Studies in Mathematical Economics and Econometrics* (1942), University of Chicago Press, 1942, pp 135-150

[18] A. C. PIGOU, "A method of determining the numerical values of elasticities of demand," *Economic Journal*, Vol. 20 (1910), pp. 636–640

[19] O. REIERSØL, "On the identifiability of parameters in Thurstone's multiple factor analysis," *Psychometrika*, Vol 15 (1950).

[20] H SCHULTZ, *Theory and Measurement of Demand*, University of Chicago Press, 1938.

[21] L L. THURSTONE, *The Vectors of Mind*, University of Chicago Press, 1935

[22] L L. THURSTONE, *Multiple-factor Analysis*, University of Chicago Press, 1947.

[23] J. TINBERGEN, *Business cycles in the United States of America, 1919–1932*, League of Nations, Geneva, 1939.

[24] A. WALD, "Note on the identification of economic relations," *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10, New York, John Wiley and Sons, 1950.

[25] S WRIGHT, "The method of path coefficients," *Annals of Math Stat.*, Vol. 5 (1934), p. 161.

[26] F. YATES, "The design and analysis of factorial experiments," *Imp. Bur. Soil Sci. Tech. Comm.*, No. 35, 1937.

[27] G. H. THOMSON, "The proof or disproof of the existence of general ability," *Brit. Jour. Psych.*, Vol. 9 (1919), p. 323.

SOME PROBLEMS IN MINIMAX POINT ESTIMATION

By J. L. Hodges, Jr.[1] and E. L. Lehmann

University of California, Berkeley

1. Summary. In the present paper the problem of point estimation is considered in terms of risk functions, without the customary restriction to unbiased estimates. It is shown that, whenever the loss is a convex function of the estimate, it suffices from the risk viewpoint to consider only nonrandomized estimates. For a number of specific problems the minimax estimates are found explicitly, using the squared error as loss. Certain minimax prediction problems are also solved.

2. Introduction. The principles most commonly applied in the selection of a point estimate are the principles of maximum likelihood (R. A. Fisher) and of minimum variance unbiased estimation (Markoff).[2] Both of these principles are intuitively appealing, but neither of them can be justified very well in a systematic development of statistics. This holds also for some modifications of these principles proposed by G. W. Brown [1], as the author himself points out.

In an important early paper [2], Wald indicated a more systematic approach to the problem, which he later developed into his general theory of statistical decision problems [3, 4, 5]. Consider a random variable X distributed over a space $\mathfrak{X}$ according to a distribution P_θ^X with $\theta \in \Omega$. It is desired to estimate some $g(\theta)$. If the value x of X is observed one makes an estimate, say $f(x)$, and thereby incurs a loss of $W[g(\theta), f(x)]$ when θ is the true value of the parameter. We shall assume that the loss function is nonnegative. It then follows that the expectation of the loss will always exist (although it may be infinite). The risk associated with the estimate f is defined to be the expected loss, as given by

$$R_f(\theta) = E_\theta W[g(\theta), f(x)] = \int_{\mathfrak{X}} W[g(\theta), f(x)]\, dP_\theta^X(x). \tag{2.1}$$

The choice of estimate should then be made according to the risk function. As a particular possibility Wald suggests the use of minimax estimates, i.e. estimates which minimize $\sup_\theta R_f(\theta)$.

The main purpose of the present paper is to obtain minimax estimates for a number of specific problems. Only few such problems have been worked out so far, the emphasis in Wald's work having been on the general theory. In [2] Wald obtained the minimax estimate of an unknown location parameter. Stein and Wald [6] treated the sequential problem of estimating the mean of a normal dis-

[1] This work was supported in part by the Office of Naval Research.

[2] Actually, the principle of minimum variance unbiased estimation goes back to Gauss. For discussions of the history of these ideas, see E. Czuber, *Theorie der Beobachtungsfehler*, Leipzig, 1891, and R. L. Plackett, "A historical note on the method of least squares", *Biometrika*, Vol. 36 (1950), p. 458.

tribution with known variance, and in his forthcoming book Wald considers as an example the sequential problem of estimating the mean of a random variable distributed uniformly over an interval of length 1.

It seems worthwhile to consider further special problems both because one may obtain estimates that in some cases are preferable to the conventional ones, and because these examples throw some light on the general desirability of the minimax principle. As we shall see below, it does not seem possible to reach any definite conclusions on this latter point, and to obtain a generally valid comparison between the minimax estimate and, for example, the unbiased estimate with uniformly smallest variance (when such an estimate exists).

Consider, for example, the problem of estimating the probability of success from a number of independent trials each of which may be a success or a failure, when the loss-function is the squared error. If the number of trials is one, the minimax estimate (as is shown below) is given by $f(X) = \frac{1}{2}X + \frac{1}{4}$, where X is 1 or 0 as the trial is a success or failure. As is easily seen, this estimate has smaller risk than the usual estimate $f^*(X) = X$ whenever $0.07 \leqq p \leqq 0.93$. On the other hand, when the number of trials is large the standard estimate $\bar{X}$ has smaller risk than the minimax estimate nearly everywhere. The minimax estimate is only slightly better in a small interval centered at $p = \frac{1}{2}$, whose length tends to zero as the number of trials tends to infinity, and is worse everywhere else.

For our purpose it is convenient to formulate the problem of point estimation as follows (see in this connection [7]). A random variable X is distributed over a space $\mathfrak{X}$ according to a distribution P belonging to a family $\mathfrak{F}$. We wish to estimate $g(P)$ where g is a function whose domain is $\mathfrak{F}$ and whose range is contained in some space $\mathfrak{A}$ (in any example $\mathfrak{A}$ is usually a Euclidean space, mostly even a one dimensional Euclidean space). An estimate is a statistic $f(X)$ taking on values in $\mathfrak{A}$. We denote by $W[g(P), f(x)]$ the loss which results from making the estimate $f(x)$ when P is the true distribution, and we define the risk function of the estimate f by

$$R_f(P) = E_P W[g(P), f(X)]. \tag{2.2}$$

The problem is to determine f so as to minimize $\sup_{P \in \mathfrak{F}} R_f(P)$.

Our principal tool will be the following theorem, which is essentially contained in Wald's work but which is not stated there explicitly. The theorem is a slight modification of one used for the theory of testing in [8].

THEOREM 2.1 *Let $\{P_\theta\}$, $\theta \in \omega$ (where ω is a subset of a Euclidean space), be a parametric subfamily of $\mathfrak{F}$, and let λ be a probability measure over ω. Suppose that f minimizes*

$$\int_\omega E_\theta W[g(P_\theta), f(X)]\, d\lambda(\theta) \tag{2.3}$$

and that

(i) *$E_\theta W[g(P_\theta), f(X)]$ is constant (say c) for all $\theta \in \omega$,*

(ii) *$E_P W[g(P), f(X)] \leq c$ for all P in $\mathfrak{F}$*

Then f is a minimax estimate for estimating g.

PROOF. Let f^* be any other estimate of g. Then

$$\sup_{P\in\mathfrak{F}} E_P W[g(P), f(X)] = \int_\omega E_\theta W[g(P_\theta), f(X)]\, d\lambda(\theta)$$

$$\leq \int_\omega E_\theta W[g(P_\theta), f^*(X)]\, d\lambda(\theta) \tag{2.4}$$

$$\leq \sup_{P\in\mathfrak{F}} E_P W[g(P), f^*(X)].$$

We note that if f is the unique function minimizing (2.3), then the first inequality in (2.4) becomes strict, and hence f is the unique minimax estimate of g.

Following Wald we shall call the function f that minimizes (2.3) the Bayes estimate of g associated with the a priori distribution λ. As a corollary to theorem 2.1, we note that a Bayes estimate whose risk function is constant, is a minimax estimate.

3. Randomization. In the formulation of the problem of point estimation given above, the estimate $f(x)$ is assumed to be completely determined by the observed value x of the random variable X. In the present section a broader formulation of the problem will be considered, in which the estimate corresponding to x may itself be a random variable, say T_x. This extension is a special case of the notion of randomized decision function introduced by Wald in his general decision theory. We associate with each x in $\mathfrak{X}$ a probability distribution F_x, with the convention that when X is observed to have the value x, we estimate $g(P)$ by means of a random variable T_x which is distributed according to F_x. Estimates of this latter kind we shall call *randomized*, and the fixed estimates $f(x)$ *nonrandomized*.

The motivation behind the admission of randomized estimates (or more generally of randomized statistical decision funtions) is that in some problems of statistical inference the performance of the decision function is considerably improved by randomization. It is clear however that the randomized functions are more complicated, and hence that it is useful to know when their consideration is not necessary. Before investigating this question we give the following definition, which makes precise a sense in which certain estimates may be omitted from consideration. (See Wald [9]).

DEFINITION For a given estimation problem a class C of estimates will be said to be essentially complete with respect to a class D of estimates, if for every estimate g in D there exists an estimate f in C such that $R_f(P) \leq R_g(P)$ for all P in $\mathfrak{F}$. If D is the class of all randomized estimates we simply say that C is essentially complete for the given problem.

It is clear that if one adopts the risk function point of view, one loses nothing by restricting consideration to an essentially complete class of estimates. In the present section we find conditions under which the totality of nonrandomized estimates forms an essentially complete class.

For this purpose we need the notion of convexity. A set S in a k-dimensional Euclidean space is said to be convex if, whenever P and Q are in S, then all points on the line segment from P to Q are also in S. A real valued function ψ defined over a k-dimensional Euclidean space is said to be convex, if for any points $(x_1, \cdots, x_k)$ and $(y_1, \cdots, y_k)$ of the space, and any number $0 < \alpha < 1$ we have

$$\alpha\psi(x_1, \cdots, x_k) + (1 - \alpha)\psi(y_1, \cdots, y_k) \geqq \psi(\alpha x_1 + (1 - \alpha)y_1, \cdots, \alpha x_k + (1 - \alpha)y_k). \tag{3.1}$$

We use the following notation for conditional expectation. If U and V are two random variables which have a joint distribution, then $E(U|v)$ denotes the conditional expectation of U given that $V = v$, $E(U|S)$ denotes the conditional expectation of U given that V is in S. Let $\Phi(v) = E(U|v)$, then for $\Phi(V)$ we write $E(U \mid V)$

LEMMA 3.1. *Let U, V be two random variables with a joint distribution, such that U is distributed in a k-dimensional space and $E(U)$ is finite. Let ψ be a real-valued convex function defined over this space and bounded from below. Then*

$$E\{\psi\{E(U \mid V)\}\} \leqq E\{\psi(U)\}.$$

PROOF. The proof is immediate in the special case that, for almost all v, there exists a determination of the conditional probability distribution of U given v which is a measure. We then know, from the convexity of ψ, that for almost all values v of V, $\psi\{E(U \mid v)\} \leqq E\{\psi(U) \mid v\}$. Replacing v by V and taking expectations of both sides, we obtain the desired result.

If we do not assume the existence of conditional measures, the proof is more complicated. Since $E(U)$ is finite, there exists a function $E(U \mid v)$ such that for any set S, $E(U \mid S) = E\{E(U \mid V) \mid S\}$; see [10], p. 47. Since ψ is convex it is measurable, and since ψ is bounded from below $E\{\psi(U)\}$ exists. Excluding the trivial case $E\{\psi(U)\} = \infty$, we know there exists a function $E\{\psi(U) \mid v\}$ such that for any set S, $E\{\psi(U) \mid S\} = E\{E\{\psi(U) \mid V\} \mid S\}$.

If the lemma were false, we should have $E\{E\{\psi(U) \mid V\}\} < E\{\psi\{E(U \mid V)\}\}$, and could find an $\epsilon > 0$ and a set A of positive V measure such that for every $v \in A$, $E\{\psi(U) \mid v\} + 2\epsilon < \psi\{E(U \mid v)\}$. This implies the existence of a number d and a set B of positive V measure such that for every $v \in B$, $E\{\psi(U) \mid v\} \leqq d$ and $d + \epsilon \leq \psi\{E(U \mid v)\}$. Since ψ is convex, the domain D of points P for which $\psi(P) < d + \epsilon$ is convex, and we may find a subset C of B, of positive V measure, for which the set of points $E(U \mid v)$, $v \in C$, lies in a convex domain E disjoint of D. It follows that $E(U \mid C)$ lies in E, and hence that $\psi\{E(U \mid C)\} \geqq d + \epsilon$. Clearly $d \geqq E\{\psi(U) \mid C\}$. Thus we have the contradiction $E\{\psi(U) \mid C\} \geq \psi\{E(U \mid C)\}$.

DEFINITION A loss function W will be called convex if for every $u \in \mathcal{U}$, $W(u, v)$ is a convex function of the estimate v.

An example of a convex loss function is provided by the Markoff principle of estimation. The variance of an unbiased estimate may be considered as a risk

function if we take the loss function to be the squared error, i.e. the square of the difference between the true value $g(P)$ and the estimated value $f(x)$ or T_x; and this loss function is clearly convex.

THEOREM 3.2. *If the loss function W is convex, if $\mathcal{G}$ is a Euclidean space, and if we consider only estimates having finite expectation, then the class of nonrandomized estimates is essentially complete.*

PROOF. Let T_X be any randomized estimate such that $E(T_X)$ exists and is finite. Applying lemma 3.1 we see that $E(T_X \mid X)$, which is a function of X only, is a nonrandomized estimate, has a risk never greater than that of T_X.

The restriction in theorem 3.2 to estimates having finite expectation may be replaced by the requirement that for each $u \in \mathcal{G}$ there exist a number M_u such that if $|v - u| = M_u$ then $W(u, v) > W(u, u)$. With this requirement and the convexity assumption, it follows that the risk associated with T_X is infinite whenever $E(T_X)$ is infinite.

Theorem 3.2 is related to a generalization of a theorem of Blackwell. If Y is a sufficient statistic for $g(P)$, and if for almost all y the conditional distribution of X given y exists in the sense of measure, we may regard estimation of $g(P)$ based on X as randomized estimation of $g(P)$ based on Y, and if the assumptions of theorem 3.1 are satisfied, we may apply this theorem to conclude the essential completeness of the class of nonrandomized estimates based on Y. In the general case we may resort again to lemma 3.1 to prove the following theorem, the proof is the same as that of theorem 3.2 if X is replaced by Y throughout.

THEOREM 3.3. *If the loss function W is convex, if $\mathcal{G}$ is in a Euclidean space, if we consider only estimates having a finite expectation, and if Y is a sufficient statistic for $\mathcal{F}$, then the class of nonrandomized estimates which are functions of Y only is essentially complete.*

Blackwell [11] proved that if U is a sufficient statistic for a real-valued parameter θ, and if T is an unbiased estimate for θ, then $E(T \mid U)$, which is a function of U only and also an unbiased estimate for θ, has a variance which never exceeds that of T. Observing that the theorems above hold true when we restrict attention to unbiased estimates, Blackwell's result may be obtained from theorem 3.3 by letting $\mathcal{G}$ be one-dimensional, letting W be the squared error, and restricting ourselves to unbiased estimates. In a similar manner we can get from theorem 3.3 an extension of Blackwell's theorem given by Barankin [12], who treated the case in which $W(\theta, t) = |\theta - t|^s$, $s \geq 1$. It is clear that these loss functions are convex.

If the convexity assumption is removed, theorems 3.2 and 3.3 cease to be true. For example, if $\mathfrak{X}$ has only n points, if $\mathcal{G}$ is a finite line segment of length greater than $2n\alpha$, and if the loss is 0 whenever $|g(P) - f(x)| \leq \alpha$, and 1 otherwise, then the minimax risk among nonrandomized estimates is 1. By admitting randomization, however, the maximum risk can be brought below 1 without using X at all; if our estimate T is uniformly distributed over $\mathcal{G}$, then the maximum risk will be $1 - \alpha/(\text{length of } \mathcal{G})$.

The example just given may seem inappropriate, in that with the specified loss

function the problem would customarily be considered one of interval estimation rather than point estimation. This objection does not apply however to the loss functions considered in the following theorem.

THEOREM 3.4. *Let* $\mathfrak{X} = \{0, 1, \cdots, n\}$, $n \geq 1$. *Let* $\mathfrak{F}$ *be the set of binomial distributions* P_p *defined by* $P_p(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$, $0 \leq p \leq 1$, *Let* $\mathfrak{D}$ *be the closed interval* $[0, 1]$ *and* $g(P_p) = p$. *Let* $W(p, t) = | p - t |^s$, $0 < s < 1$. *Then no minimax estimate can be nonrandomized, and the class of nonrandomized estimates is not essentially complete.*

PROOF. For any nonrandomized estimate f, $R_f(p)$, being a sum of products of continuous functions of p, is itself a continuous function of p. The nonrandomized minimax risk is less than 1, as may be shown by considering any estimate of the following kind: $f(0) = 0$, $f(n) = 1$, and $0 \leq f(x) \leq 1$ for all x. Here $R_f(0) = R_f(1) = 0$, while if $0 < p < 1$, $R_f(p) \leq \max_x | p - f(x) |^s < 1$. By continuity $\sup_{0 \leq p \leq 1} R_f(p) < 1$.

It is easy to see that there exists among the nonrandomized estimates a minimax estimate, say h. Let the corresponding minimax risk be denoted by M. We know that $M = \sup_{0 \leq p \leq 1} R_h(p) < 1$, it is obvious that $M > 0$. Observe that $h(0) < 1$, since $h(0) \geq 1$ leads to the contradiction $R_h(0) = | h(0) |^s \geq 1$. We can write

$$R_h(p) = \sum_{h(x) = h(0)} P_p(X = x) \cdot | p - h(x) |^s + \sum_{h(x) \neq h(0)} P_p(X = x) \cdot | p - h(x) |^s.$$

The second sum has a finite derivative with respect to p at $p = h(0)$, while the first sum increases with infinite speed as p is moved away from $h(0)$. Therefore $R_h\{h(0)\} < M$; and by an exactly symmetrical argument, $0 < h(n)$ and $R_h\{h(n)\} < M$. Using the continuity of R_h, we can find a positive number ω so small that $R_h(p) < M$ whenever $| p - h(0) | < \omega$ or $| p - h(n) | < \omega$.

Consider now the randomized estimate T_X defined by $T_x = h(x)$ if $0 < x < n$, and by $T_x = h(x) + \alpha Y$ otherwise, where Y is a random variable independent of X and taking on the values 1 and -1 each with probability $\frac{1}{2}$, and where $0 < \alpha < \omega$. Observe

$$R_{T_X}(p) - R_h(p) = (1 - p)^n [\tfrac{1}{2}\{ | p - h(0) + \alpha |^s + | p - h(0) - \alpha |^s \} - | p - h(0) |^s] + p^n [\tfrac{1}{2}\{ | p - h(n) + \alpha |^s + | p - h(n) - \alpha |^s \} - | p - h(n) |^s].$$

By the concavity of the functions involved, the first square bracketted term is negative whenever $| p - h(0) | \geq \alpha$, and the second is negative whenever $| p - h(n) | \geq \alpha$. We can choose α so small that whenever either $| p - h(0) |$ or $| p - h(n) |$ is less than α, $R_{T_X}(p) - R_h(p) < \omega$. A continuity argument now shows that $\sup_{0 \leq p \leq 1} R_{T_X}(p) < M$. But this proves that no minimax estimate, with randomization permitted, can be nonrandomized. It is also now obvious that the class of nonrandomized estimates is not essentially complete: every nonrandomized estimate must have a risk function which somewhere exceeds $\sup_{0 \leq p \leq 1} R_{T_X}(p)$.

4. General properties of minimax estimation. Whether a principle such as the minimax principle is a desirable one has to be decided mainly on two criteria:

(i) its general properties, and

(ii) its performance in many particular instances.

It has already been remarked that in the second respect the minimax principle does not seem entirely satisfactory. With regard to the former, one great advantage of this principle is that when there is a unique minimax estimate, it is admissible. Here an estimate f is said to be admissible (see [3]) if there exists no other estimate f^* such that $R_{f^*}(P) \leq R_f(P)$ for all P in $\mathcal{F}$ with strict inequality holding for some P. It is interesting that, as we shall show below, this admissibility property is not shared by either the principle of unbiasedness or the maximum likelihood principle.

In this connection we begin by proving another theorem concerning essentially complete classes.

THEOREM 4.1. *Suppose that the space $\mathcal{Y}$ is a finite interval $[a, b]$ on the real line, and that for each $u \in \mathcal{Y}$, $W(u, v)$ is a non-decreasing function of v when $v > u$ and a non-increasing function of v when $v < u$. Then the class of estimates whose range is contained in $\mathcal{Y}$ is essentially complete with respect to the class of all real valued estimates.*

PROOF. If T is any real-valued estimate, define T^* by

$$T^* = \begin{cases} T & \text{if} \quad T \in \mathcal{Y}, \\ a & \text{if} \quad T < a, \\ b & \text{if} \quad t > b. \end{cases} \tag{4.1}$$

It is clear that $R_{T^*}(P) \leq R_T(P)$ for every $P \in \mathcal{F}$.

Halmos [7] has provided an example in which the minimum variance unbiased estimate takes on, with positive probability, values outside the range of the parameter. It can be shown from the proof of theorem 4.1 that in this case any unbiased estimate is inadmissible, provided the loss function is of the kind described in theorem 4.1.

That the maximum likelihood principle may also lead to inadmissible estimates is easy to show, since this is the case in many familiar situations. The following example may be of interest in that here the maximum likelihood estimate is uniformly worst among all estimates which one would consider using.

Example Let X be a random variable with only 0 and 1 as possible values, and let $P(X = 1) = p$. Assume it to be known that $\frac{1}{3} \leq p \leq \frac{2}{3}$. Then the maximum likelihood estimate for p is easily seen to be $\frac{1}{3}(X + 1)$, and, if the loss function is the squared error, the associated risk function is $\frac{1}{3}(p - \frac{1}{2})^2 + \frac{1}{36}$. This risk function is, for every possible value of p, greater than that of any estimate $f(x)$ satisfying $\frac{1}{3} \leq f(0) \leq f(1) = 1 - f(0) \leq \frac{2}{3}$.

The selection of loss function in any problem should in theory be governed by metastatistical considerations, but in fact the circumstances of statistical problems do not usually offer compelling reasons for using one loss function rather

than another. Considerations of mathematical facility are often determining. Thus, various classical unbiased estimates become minimax estimates when the loss function is judiciously chosen. For, if we take as loss function the ratio of squared error to the variance of the unbiased estimate, the risk becomes constant, and we can easily obtain the classical estimates as minimax estimates in the familiar binomial, Poisson, and rectangular problems, and in some of the non-parametric problems considered in section 6.

However, this approach seems to be somewhat artificial, and hereafter we shall restrict ourselves to a single loss function, namely the squared error. There are two reasons for this choice. With squared error for the loss, the mathematical problems are rather simple. And as was remarked above, squared error (if one restricts oneself to unbiased estimates) is the traditional loss function. Fortunately, the squared error loss function is convex, and hence theorem 3.2 permits us to avoid considering randomized estimates.

When the loss function is squared error, we have the following obvious linearity property, which for later reference we state as

THEOREM 4.2. *If $f(X)$ is the minimax estimate for $g(P)$, then $af(X) + b$ is the minimax estimate for $a \cdot g(P) + b$.*

However, as we shall show by an example in the next section, it need not be true that if $X_1, \cdots, X_n$ are independent and $f_i(X_i)$ is the minimax estimate for $g_i(P_i)$, $i = 1, \cdots, n$, then $\sum_{i=1}^n a_i f_i(X_i)$ is the minimax estimate for $\sum_{i=1}^n a_i g_i(P_i)$. This is a definite disadvantage of the minimax principle as compared with the Markoff principle which does possess the linearity property mentioned.

We conclude this section with an explicit solution of the Bayes problem in the squared error case. If the distribution P is itself a random variable distributed over $\mathcal{F}$ according to some distribution λ, we may compare estimates f by means of their expected loss $Q(f) = E[g(P) - f(X)]^2$. Since $Q(f) = E\{E[g(P) - f(X)]^2 \mid X\}$, it is well known that $Q(f)$ is minimized by using the estimate $f(x) = E[g(P) \mid x]$, provided the conditional measures exist. In fact, this result holds even without this assumption.

THEOREM 4.3. *$E[g(P) - f(X)]^2$ is minimized by $f(x) = E[g(P) \mid x]$.*

PROOF.
$$E[g(P) - f(X)]^2 - E\{g(P) - E[g(P) \mid X]\}^2 = E\{E[g(P) \mid X] - f(X)\}^2$$
$$+ 2E[E\{g(P) - E[g(P) \mid X]\}\{E[g(P) \mid X] - f(X)\} \mid X] \geq 0.$$

In applications it is convenient to write $E[g(P) \mid X]$ more explicitly. Suppose that with respect to some measure μ over $\mathfrak{X}$, each distribution $P \in \mathcal{F}$ has a generalized probability density p_P, so that for any A, the probability that $X \in A$ computed for P, is given by

$$\int_A p_P(x)\, d\mu(x)\,.$$

Minimizing a quadratic expression shows that

$$(4.2) \qquad \frac{\int_\Omega g(P) p_P(x)\, d\lambda(P)}{\int_\Omega p_P(x)\, d\lambda(P)}$$

is a Bayes solution.

5. Binomial and hypergeometric distributions. In the present section we shall consider three discrete minimax problems.

PROBLEM 1. (Binomial.) Let X be a binomial random variable with parameter p, $0 \leq p \leq 1$, so that $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$. We shall show that the minimax estimate for p is

$$(5.1) \qquad \frac{X}{n} \cdot \frac{\sqrt{n}}{(\sqrt{n} + 1)} + \frac{1}{2(\sqrt{n} + 1)}.$$

Consider any linear estimate $\alpha X + \beta$. The risk $E_p(\alpha X + \beta - p)^2$ is a quadratic function of p which is constantly equal to β^2 when $\alpha = \frac{1}{\sqrt{n}(1 + \sqrt{n})}$ and $\beta = \frac{1}{2(1 + \sqrt{n})}$. Hence (5.1) is a constant risk estimate of p. Since it is easily seen that

$$\frac{\int_0^1 p \cdot p^k q^{n-k} \cdot p^{a-1} q^{b-1}\, dp}{\int_0^1 p^k q^{n-k} \cdot p^{a-1} q^{b-1}\, dp} = \frac{a + k}{a + b + n}, \qquad (q = 1 - p),$$

it follows that (5.1) is the Bayes estimate when p is distributed with probability density $C(pq)^{(\sqrt{n}/2)-1}$, and hence by Theorem 2.1 we conclude that (5.1) is the minimax estimate of p.

After obtaining this result we were informed that it had been obtained earlier by H. Rubin, to whom, therefore, the priority belongs.

It is interesting to compare the risk of the above estimate with that of the standard unbiased estimate X/n. We have

$$E\left(\frac{X}{n} - p\right)^2 = \frac{pq}{n},$$

$$E\left[\frac{1}{1 + \sqrt{n}}\left(\frac{X}{\sqrt{n}} + \frac{1}{2}\right) - p\right]^2 = \frac{1}{4(1 + \sqrt{n})^2}.$$

As is easily seen, $\frac{pq}{n} \leq \frac{1}{4(1 + \sqrt{n})^2}$ if and only if

$$\left| p - \frac{1}{2} \right| \geq \frac{\sqrt{1 + 2\sqrt{n}}}{2(1 + \sqrt{n})}.$$

Thus the standard estimate is better than the minimax estimate outside an interval around $p = \frac{1}{2}$ whose length decreases with increasing n, tending to 0 as n tends to infinity. However, for very small values of n the minimax estimate has the smaller risk over nearly the whole range.

PROBLEM 2. (Difference of binomials.) Let X and Y be independent binomial random variables, where $P(X = k) = \binom{n}{k} p_1^k (1 - p_1)^{n-k}$ and $P(Y = l) = \binom{n}{l} p_2^l(1 - p_2)^{n-l}$. By use of theorem 2.1 we shall show that the minimax estimate for $p_1 - p_2$ is $\frac{\sqrt{2n}}{1 + \sqrt{2n}} \left(\frac{X}{n} - \frac{Y}{n}\right)$. For the set ω of theorem 2.1 we take $p_1 = p$, $p_2 = 1 - p$, $0 \leq p \leq 1$, and we let $Z = X + n - Y$. Applying the result of Problem 1 to Z, we find the minimax estimate of p to be $\alpha_{2n} \cdot Z + \beta_{2n}$, and by Theorem 4.2 the minimax estimate based on Z for $p_1 - p_2 = 2p - 1$, is $\frac{\sqrt{2n}}{1 + \sqrt{2n}} \left(\frac{X}{n} - \frac{Y}{n}\right)$, and the risk of this estimate is constant over ω.

To prove that this is also the minimax estimate of $p_1 - p_2$ for the original problem, we consider the risk as a function of p_1 and p_2. It is easy to show that $(1 + \sqrt{2n})^2 R(p_1, p_2) = 2 \cdot [p_1(1 - p_1) + p_2(1 - p_2)] + (p_1 - p_2)^2$. Finally it can be shown that $p_1(1 - p_1) + p_2(1 - p_2)$ is maximized, subject to the condition that $p_1 - p_2$ be constant, when $p_1 + p_2 = 1$.

PROBLEM 3. (Hypergeometric.) We finally consider the problem of estimating the number of defectives in a lot from a sample drawn from this lot at random. We denote by N and n the number of elements in lot and sample respectively, and by D and X the corresponding number of defectives. For later reference we note

$$P(X = k) = \frac{\binom{D}{k}\binom{n - D}{n - k}}{\binom{N}{n}},$$

$$E(X) = n\frac{D}{N},$$

$$\sigma_x^2 = \frac{nD(N - n)(N - D)}{N^2(N - 1)}.$$

As in Problem 1 we easily find a linear function of X whose risk is constant. In fact

$$E_D(\alpha X + \beta - D)^2 \equiv \beta^2$$

when

$$\alpha = \frac{N}{n + \sqrt{\frac{n(N - n)}{N - 1}}}, \beta = \frac{N}{2}\left(1 - \frac{\alpha n}{N}\right).$$

To prove that $\alpha X + \beta$ is the minimax estimate of D we shall show that it is the Bayes estimate corresponding to

$$P(D = d) = \int_0^1 \binom{N}{d} p^d q^{N-d} \cdot C\, p^{a-1} q^{b-1}\, dp, \tag{5.2}$$

where $a, b > 0$, and

$$C = \frac{\Gamma(a + b)}{\Gamma(a)\, \Gamma(b)}.$$

In this connection it is useful to notice that since (5.2) is a distribution

$$\sum_{d=0}^{N} \binom{N}{d} \frac{\Gamma(a + d)\, \Gamma(N + b - d)}{\Gamma(N + a + b)} \cdot \frac{\Gamma(a + b)}{\Gamma(a)\, \Gamma(b)} = 1. \tag{5.3}$$

Using theorem 4 3, we find the Bayes estimate associated with (5.2) to be

$$f(k) = \frac{\sum_{d=k}^{N-n+k} d \binom{d}{k} \binom{N - d}{n - k} \binom{N}{d} \Gamma(a + d)\, \Gamma(N + b - d)}{\sum_{d=k}^{N-n+k} \binom{d}{k} \binom{N - d}{n - k} \binom{N}{d} \Gamma(a + d)\, \Gamma(N + b - d)}.$$

Replacing d by $(d - a) + a$, and using the relation

$$\binom{d}{k} \binom{N - d}{n - k} \binom{N}{d} = \binom{N - n}{d - k} \cdot \text{(terms not involving } d\text{)},$$

we find:

$$f(k) = \frac{\sum_{i=0}^{N-n} \binom{N - n}{i} \Gamma(d + a + 1)\, \Gamma(N + b - d)}{\sum_{i=0}^{N-n} \binom{N - n}{i} \Gamma(d + a)\, \Gamma(N + b - d)} - a.$$

Now apply (5.3) to numerator and denominator separately; then

$$f(k) = k\,\frac{a + b + N}{a + b + n} + \frac{a(N - n)}{a + b + n}.$$

Putting $\dfrac{a + b + N}{a + b + n} = \alpha$, $\dfrac{a(N - n)}{a + b + n} = \beta$ one obtains easily

$$a = \frac{\beta}{\alpha - 1}, \quad b = \frac{N - \alpha n - \beta}{\alpha - 1}.$$

Substituting the values of α and β one finds that $\beta > 0$, $N > \alpha n + \beta$ and that $\alpha > 1$ provided $N > n + 1$ In the special case $N = n$ the result is immediate, while if $N = n + 1$, the result is obtained by giving to D a binomial distribution with $p = \frac{1}{2}$.

6. Non parametric problems. We shall in this section consider estimation problems in which the functional form of the distribution of X is not assumed known. Restrictions will be imposed on the variables only to insure the existence of estimates with bounded risk. The problem will be treated under two different such restrictions: (i) that the variables are bounded with known bounds, (ii) that the variables have bounded variances.

In the first of these cases we can assume without loss of generality that the variables are distributed over the interval [0, 1], and then obtain

THEOREM 6.1. *Let $X_1, \cdots, X_n$ be independently distributed over* [0, 1] *according to a joint distribution belonging to a family $\mathcal{F}$. Suppose that $\mathcal{F}$ contains the subfamily $\mathcal{F}_0$ according to which $X_1, \cdots, X_n$ are independently and identically distributed with $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$, $0 \leq p \leq 1$. Let $E(X_i) = \mu_i$, $\frac{1}{n}\sum_{i=1}^{n} \mu_i = \bar{\mu}$. Then the minimax estimate of $\bar{\mu}$ is*

$$\frac{1}{1 + \sqrt{n}}\left(\sqrt{n}\,\bar{X} + \tfrac{1}{2}\right). \tag{6.1}$$

PROOF. Since (6.1) is the minimax estimate of $\bar{\mu} = p$ when the distribution of the X's is known to belong to $\mathcal{F}_0$, we only need to show that its risk is largest for the distributions of $\mathcal{F}_0$. But

$$E(A\bar{X} + B - \bar{\mu})^2 = A^2\sigma_{\bar{x}}^2 + [B + (A - 1)\bar{\mu}]^2 = \frac{A^2}{n^2}\sum_{i=1}^{n}\sigma_{x_i}^2 + [B + (A - 1)\bar{\mu}]^2$$

and

$$\Sigma\sigma_{x_i}^2 = \Sigma E(X_i^2) - \Sigma\mu_i^2 \leqq \Sigma\mu_i - \Sigma\mu_i^2 = n\bar{\mu} - \Sigma(\mu_i - \bar{\mu})^2 - n\bar{\mu}^2 \leqq n\bar{\mu}(1 - \bar{\mu})$$

where equality holds for the distributions in $\mathcal{F}_0$.

COROLLARY 6.2. *Let $X_1, \cdots, X_n$ be a sample from an unknown univariate distribution over* [0, 1]. *Then the minimax estimate of $E(X_i) = \mu$ is given by* (6.1).

COROLLARY 6.3. *Let $X_1, \cdots, X_n$ be a sample from an unknown absolutely continuous univariate distribution over* [0, 1]. *Then the minimax estimate of $E(X_i) = \mu$ is given by* (6.1).

Corollary 6.3 follows from the fact that any risk function that can be obtained for binomial distribution can be approximated by means of absolutely continuous distributions.

Theorem 6.1 can be extended to include variables that are negatively correlated. Namely if $X_1, \cdots, X_n$ are distributed over [0, 1] according to a joint distribution belonging to some family $\mathcal{F}$, if for each distribution of $\mathcal{F}$ the correlation coefficient ρ_{ij} of X_i, X_j is $\leqq 0$ for all i, j, and if $\mathcal{F}$ contains the family $\mathcal{F}_0$ of theorem 6.1, then the conclusion of this theorem remains valid. This result can be used for example in the following situation. Suppose a sample of n is taken from a lot of unknown size, and suppose it is desired to estimate the proportion p of defectives in the lot. If k is the number of defectives in the sample, it follows from the above remarks that the minimax estimate of p is $\dfrac{1}{1 + \sqrt{n}}\left(\dfrac{k}{\sqrt{n}} + \dfrac{1}{2}\right)$.

It should be pointed out that this result holds only if no upper bound is assumed known for the lot size. If it is known that the number of items in the lot is $\le N_0$, then the minimax estimate is that found in section 5 for the case of a hypergeometric distribution with $N = N_0$.

Next let us consider estimating the difference of the average means in two groups of variables.

THEOREM 6.4. *Let $X_1, \cdots, X_n; Y_1, \cdots, Y_n$ be independently distributed over the interval* [0, 1] *according to a joint distribution belonging to a family $\mathcal{F}$. Suppose that $\mathcal{F}$ contains the subfamily $\mathcal{F}_1$, according to which $X_1, \cdots, X_n$; $Y_1, \cdots, Y_n$ are two samples with $P(X_i = 1) = p_1$, $P(X_i = 0) = 1 - p_1$; $P(Y_i = 1) = p_2$, $P(Y_i = 0) = 1 - p_2$, $0 \leq p_1, p_2 \leq 1$. If $E(X_i) = \mu_i$, $E(Y_i) = \nu_i$, $\frac{1}{n}\sum \mu_i = \bar{\mu}$, $\frac{1}{n}\sum \nu_i = \bar{\nu}$, then the minimax estimate of $\bar{\mu} - \bar{\nu}$ is*

$$\frac{\sqrt{2n}}{1 + \sqrt{2n}}(\bar{X} - \bar{Y}). \tag{6.2}$$

PROOF. Again, since (6.2) is the minimax estimate in the binomial case (Problem 2 of section 5), we need only verify that its risk takes its maximum in $\mathcal{F}_1$. But

$$\begin{aligned} &E[A(\bar{X} - \bar{Y}) - (\bar{\mu} - \bar{\nu})]^2 \\ &= E[A(\bar{X} - \bar{\mu}) - A(\bar{Y} - \bar{\nu}) + (A - 1)(\bar{\mu} - \bar{\nu})]^2 \\ &= A^2(\sigma_{\bar{X}}^2 + \sigma_{\bar{Y}}^2) + (A - 1)^2(\bar{\mu} - \bar{\nu})^2, \end{aligned}$$

of which we already have shown that it is maximized in the binomial case.

Up to now we assumed the variables to be bounded. Let us now suppose instead that the variances are bounded. With this assumption we can give an analogue of the classical Markoff theorem on least squares.

THEOREM 6.5. *Suppose that $X_1, \cdots, X_n$ are independently distributed according to a joint distribution belonging to some family $\mathcal{F}$, which contains the subfamily $\mathcal{F}_0$ where the X's are normal with variance M^2. Suppose that for all distributions in $\mathcal{F}$, $E(X_i) = \sum_{j=1}^{s} a_{ij}\theta_j$, and $\sigma_{X_i}^2 \leq M^2$. We assume the matrix (a_{ij}) to be known and of rank $s \leq n$. Then the estimate $[f_1(X), \cdots, f_s(X)]$ of $(\theta_1, \cdots, \theta_s)$ which minimizes $\sup E \sum_i [f_i(X) - \theta_i]^2$, is the Markoff estimate.*

PROOF. Consider first the subfamily $\mathcal{F}_0$. Then there exists an orthogonal transformation to $Y_1, \cdots, Y_n$ such that $E(Y_i) = k_i\theta_i$ for $i = 1, \cdots, s$, where $k_i > 0$; $E(Y_i) = 0$ for $i = s + 1, \cdots, n$, and $\sigma_{Y_i}^2 = M^2$ for $i = 1, \cdots, n$. Then $(Y_1, \cdots, Y_s)$ is a sufficient statistic for $(\theta_1, \cdots, \theta_s)$, and it is easily shown, using the methods of [6], that $\left(\frac{Y_1}{k_1}, \cdots, \frac{Y_s}{k_s}\right)$ is the minimax estimate for $(\theta_1, \cdots, \theta_s)$. But this is the Markoff estimate. In order to complete the proof we must show that the risk of this estimate takes on in $\mathcal{F}_0$ its supremum over $\mathcal{F}$. But this is immediate, for $E\sum_{i=1}^{s}[f_i(X) - \theta_i]^2 = E\sum_{i=1}^{s}\left(\frac{Y_i}{k_i} - \theta_i\right)^2 \leq M^2\sum_{i=1}^{s}\frac{1}{k_i^2}$.

In a similar manner it is easily shown that the least squares estimate for a linear function of one or more of the θ's, is the minimax estimate.

Theorem 6.5 gives a justification of the least squares estimate different from that of the Markoff theorem. In the Markoff theorem, it is shown that the least squares estimate has *uniformly* smallest risk among all linear unbiased estimates, here it is shown that the least squares estimate minimizes the maximum risk among all estimates. (The assumptions concerning variances also differ.)

7. Prediction problems. Frequently one is interested in estimating the value of a random variable rather than that of a parameter. A customary method for this is to estimate the expectation of the random variable (a parameter) and then to "identify" the variable and its expectation; i.e., to use the estimate of the expectation as a prediction for the variable. As we shall see below one is led to this procedure if one adopts the point of view of unbiased estimation, so that from this point of view prediction poses no new problem. This however is no longer true when one employs the minimax principle.

Consider a pair X, Y of random variables having a joint distribution P belonging to a family $\mathfrak{F}$ of distributions. It is desired to use the observed X to predict, say, $g(Y)$. We are interested in minimax predictions, i.e., functions $f(X)$ which minimize $\sup_{P \epsilon \mathfrak{F}} E_P W[g(Y), f(X)]$. To obtain minimax predictions we need the following analogue of Theorem 2.1.

THEOREM 7.1. *Let $\{P_\theta\}$, $\theta \epsilon \omega$ be a parametric subfamily of $\mathfrak{F}$, and let λ be a probability measure over ω. Suppose that f is such that $\int E_\theta W[g(Y), f(X)]\, d\lambda(\theta)$ is minimum, and that*

(i) $E_\theta W[g(Y), f(X)]$ is constant, say $= c$, for all $\theta \epsilon \omega$,

(ii) $E_P W[g(Y), f(X)] \leq c$ for all $P \epsilon \mathfrak{F}$.

Then f is a minimax prediction for $g(Y)$.

The proof is the exact analogue of that of theorem 2.1.

COROLLARY 7.2. *A constant risk Bayes prediction is a minimax prediction.*

Suppose now that X and Y are independent and that $W[g(y), f(x)] = [g(y) - f(x)]^2$. Consider the problem first from the point of view of unbiasedness. A prediction could reasonably be called unbiased if $E_P f(X) = E_P g(Y)$. Subject to unbiasedness, the risk is given by $E_P[g(Y) - f(X)]^2 = \sigma_P^2 f(X) + \sigma_P^2\, g(Y)$. But $\sigma_P^2 g(Y)$ is a known function of P, and hence the problem of minimizing (for a particular P) the expected squared error reduces to that of finding an unbiased estimate of $E_P g(Y)$ with minimum variance at P. In a similar way one sees, without any restriction to unbiased predictions, that the Bayes prediction for $g(Y)$ is the same as the Bayes estimate for $E_P g(Y)$, and hence that formula (4.2), with $g(P)$ replaced by $E_P g(Y)$, may be used if the assumptions there made are valid.

One might expect that as in the unbiased theory the prediction will coincide with the estimate. This however is not the case since the λ's that give constant risk in the two cases will usually be distinct. In fact the two problems are rather

different in that the "least favorable" λ for the prediction problem must not only take into account the difficulty of finding the correct value of θ for various a priori distributions but also the difficulty of predicting $g(Y)$ when θ is known.

As a first example consider the prediction analogue of problem 1 of section 5. Let X, Y be independent binomial variables such that $P(X = k) = \binom{m}{k} p^k(1-p)^{m-k}$ and $P(Y = l) = \binom{n}{l} p^l(1-p)^{n-l}$. We shall obtain the minimax prediction of Y in a manner quite analogous to the one in which we determined the minimax estimate of p. Actually, the present problem is a generalization of the earlier one, to which it can be reduced by letting $n \to \infty$. First it is easily seen that

$$E\left(\alpha \frac{X}{m} + \beta - \frac{Y}{n}\right)^2$$

is a quadratic function of p, which when $m > 1$ is constant for

$$\alpha = \frac{m}{m-1}\left[1 - \sqrt{\frac{1}{m} + \frac{1}{n} - \frac{1}{mn}}\right],$$

$$\beta = \frac{1-\alpha}{2}.$$

But we have already seen that $\alpha \frac{X}{m} + \beta$ is the Bayes solution corresponding to $Cp^{a-1}q^{b-1}$ where $\alpha = \frac{m}{m+a+b}$, $\beta = \frac{a}{m+a+b}$. Clearly $\beta = \frac{1-\alpha}{2}$ when $a = b$, and $a > 0$ provided $0 < \alpha < 1$, which is easily verified when $m, n > 1$. We note that as $n \to \infty$, the values of α, β tend to those of the minimax estimate of P.

When $m = 1$, $E\left(\alpha \frac{X}{m} + \beta - \frac{Y}{n}\right)^2$ is constant for $\alpha = \frac{n-1}{2n}$, $\beta = \frac{1-\alpha}{2}$, and again $\alpha \frac{X}{m} + \beta$ is the Bayes estimate of a beta distribution when $n > 1$, and hence minimax.

Finally in the case $n = 1$, the situation degenerates. Since $E(\frac{1}{2} - Y)^2 = \frac{1}{4}$, the prediction $f(X) = \frac{1}{2}$ has constant risk. In addition it is the Bayes prediction corresponding to the distribution which assigns probability 1 to $p = \frac{1}{2}$. Hence in this case, regardless of the value of X one would predict for Y the value $\frac{1}{2}$.

It is interesting that the above prediction problem can be interpreted also as an estimation problem in the following manner. Suppose a lot of size $N = m + n$ is such that the number of defectives follow a binomial distribution; this is the case when the items making up the lot are produced by a manufacturing process that is in statistical control. It is desired to estimate from a sample of size m taken from this lot, the proportion of defectives in the remainder. That this is equivalent to the prediction problem treated above follows from a remark of Mood [13] that in such a lot the number of defectives in the sample and in the remainder are independently distributed according binomial distributions with common p.

We can again use the binomial results to obtain the solutions of certain non-parametric problems. For example, let $X_1, \cdots, X_m$ be independently and identically distributed on [0, 1] and let $Y_1, \cdots, Y_n$ be another sample from the same distribution. Then the minimax prediction for $\bar{Y}$ is given by $\alpha\bar{X} + \beta$ with $\alpha = \frac{m}{m-1}\left[1 - \sqrt{\frac{1}{m} + \frac{1}{n} - \frac{1}{mn}}\right]$, $\beta = \frac{1-\alpha}{2}$. This follows from the fact that

$$
\begin{aligned}
E(\alpha\bar{X} + \beta - \bar{Y})^2 &= E[\alpha(\bar{X} - \bar{\mu}) - (\bar{Y} - \mu) + (\beta + (\alpha - 1)\mu)]^2 \\
&= \alpha^2\left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2 + [\beta + (\alpha - 1)\mu]^2 \\
&\leq \alpha^2\left(\frac{1}{m} + \frac{1}{n}\right)\mu(1 - \mu) + [\beta + (\alpha - 1)\mu]^2.
\end{aligned}
$$

An analogous modification clearly is possible for theorem 6.4.

For the situation considered in 6.5, the prediction problem gives the same result as the estimation problem. For consider first two samples $X_1, \cdots, X_m$; $Y_1, \cdots, Y_n$ from a normal distribution with known variance σ^2. Here

$$
E_\theta[f(X_1, \cdots, X_n) - \bar{Y}]^2 = E_\theta[f(X_1, \cdots, X_m) - \theta]^2 + \frac{\sigma^2}{n},
$$

and hence the risk differs from that of the estimation problem only by a constant. Thus $\bar{X}$ is the minimax prediction of $\bar{Y}$, and it is then seen immediately that it is also the minimax prediction for $\bar{Y}$ when of the underlying common distribution of the X's and Y's it is assumed only that the variance is bounded.

REFERENCES

[1] G. W. Brown, "On small sample estimation," *Annals of Math. Stat.*, Vol. 18 (1949), p. 514.

[2] A. Wald, "Contributions to the theory of statistical estimation and testing hypotheses," *Annals of Math. Stat.*, Vol. 10 (1939), p. 299.

[3] A. Wald, *On the Principles of Statistical Inference*, Notre Dame Math. Lectures, No. 1 (1942).

[4] A. Wald, "Statistical decision functions which minimize the maximum risk," *Annals of Math*, Vol. 46 (1945), p. 265

[5] A. Wald, "Statistical decision functions," *Annals of Math. Stat*, Vol. 20 (1949), p 165.

[6] C Stein and A Wald, "Sequential confidence intervals for the mean of a normal distribution with known variance," *Annals of Math Stat.*, Vol. 18 (1947), p 427

[7] P. R. Halmos, "The theory of unbiased estimation," *Annals of Math Stat.*, Vol. 17 (1946), p. 34.

[8] E. L. Lehmann and C. Stein, "Most powerful tests of composite hypotheses I. Normal distributions," *Annals of Math. Stat.*, Vol. 19 (1948), p 495.

[9] A. Wald, "An essentially complete class of admissible decision functions," *Annals of Math. Stat.*, Vol. 18 (1947), p. 549.

[10] A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin, 1933

[11] D Blackwell, "Conditional expectation and unbiased sequential estimation," *Annals of Math. Stat.*, Vol 18 (1947), p. 105

[12] E W Barankin, "Extension of a theorem of Blackwell," *Annals of Math. Stat*, Vol 21 (1950), p 280.

[13] A M. Mood, "On the dependence of sampling inspection plans upon population distributions," *Annals of Math Stat.*, Vol 14 (1943), p. 145.

THE THEORY OF PROBABILITY DISTRIBUTIONS OF POINTS ON A LATTICE[1]

By P. V. Krishna Iyer

University of Oxford

1. Introduction and summary. This paper discusses the theory of certain probability distributions arising from points arranged in the form of lattices in two, three and higher dimensions. The points are of k characters which for convenience are described as colors. A two dimensional lattice will consist of $m \times n$ points in m columns and n rows. In a three dimensional lattice there will be $l \times m \times n$ points in the form of a rectangular parallelopiped. Two situations arise for consideration. They are, to use the term of Mahalanobis, *free* and *non-free* sampling. In free sampling the color of each point is determined, on null hypothesis, independently of the color of the other points. The probabilities of the points belonging to the different colors, say black, white, etc. are $p_1, p_2 \cdots p_k$, such that $\sum_1^k p_r = 1$. In non-free sampling the number of points of each color is specified in advance, say $n_1, n_2 \cdots n_k$ so that $\sum_1^k n_r = mn$ or lmn according as the lattice is two- or three-dimensional. Only the arrangements of these points in the lattice are varied.

The distributions considered in this paper are the following:

(i) the number of joins between adjacent points of the same color, say black-black joins,

(ii) the number of joins between adjacent points of two specified colors, say black-white joins, and

(iii) the total number of joins between points of different colors, along mutually perpendicular axes.

The methods used here are the same as those developed by the author [3] for the linear case. All the distributions tend to the normal form when l, m and n tend to infinity, provided the p's are not very small.

Before considering the various distributions, we shall have a brief review of the work done on this topic by other people. For free sampling, Moran [5] and [6] has discussed the distribution of black-white and black-black joins for an $m \times n$ lattice of points of two colors. For a three-dimensional lattice, he has given the first and the second moments for the distribution of black-white joins. Levene [4] has announced some results closely allied to those of Moran for a square of side N (with N^2 cells) each cell taking the characteristic A or B with probabilities p and $q = 1 - p$ respectively. Bose [2] has found the expectation of

x = the number of black patches − the number of embedded white patches,

[1] Part of a thesis approved for the degree of Doctor of Philosophy, Oxford University.

for a square divided into n^2 small cells, having p and $q = 1 - p$ as the probability of the cells being black or white. An embedded white patch is one that lies completely inside a black patch.

The above review shows that the work done so far is confined entirely to the free sampling distributions, the points taking only two characters. As mentioned in the beginning of this article, we shall deal here with the free and non-free sampling distributions for points possessing k characters or colors.

2. Two dimensional lattice. Let an $m \times n$ rectangular lattice consist of mn points of k colors with probabilities $p_1, p_2, \cdots p_k$, such that $\Sigma p_r = 1$ (When there are only two colors, p_1 and p_2 are taken as p and q respectively.) All the problems dealt with for the linear lattice (Krishna Iyer, [3]) can be investigated here also. But the most important of them is the distribution for the total number of joins between points of different colors. This takes into consideration the relative position of points of all colors in the lattice. Distributions for the number of black-black or black-white joins are not based on the arrangement of all the points in the lattice and therefore cannot be considered to be adequate for testing the random distribution of the points in the lattice. Therefore the distribution of the total number of joins between points of different colors has been dealt with in some detail. As the actual distributions are very complicated they are discussed by means of cumulants. The first and the second moments for the other distributions have also been given.

2.1. *First and second moments for the distribution of black-black joins for two or more colors.* The first and the second moments for free sampling have been obtained by Moran [5] and [6]. In order to give an idea of the methods used in this paper for obtaining the moments and also to facilitate the derivation of the corresponding moments for non-free sampling, they have been obtained again for both black-black and black-white joins

(a) *Free Sampling.* In the course of similar investigations on the distribution of black-black joins arising from points on a line, the author [3] has found that the rth factorial moment is $r!$ times the sum of expectations of the different ways of obtaining r joins. This finding is true for the rectangular lattice also This may be established as follows.

Define variates $u_{ij'}$ ($i = 1, 2, \cdots, n$; $j' = 1, 2, \cdots, m - 1$) to be one if the (i, j) and $(i, j + 1)$ positions are black and zero otherwise; then $E(u_{ij'}) = p^2$, and the higher factorial moments are zero. Similarly, define $v_{i'j}$ ($i' = 1, 2, \cdots, n - 1$; $j = 1, 2, \cdots, m$) to be one when the (i, j) and $(i + 1, j)$ positions are black and zero otherwise; then $E(v_{i'j}) = p^2$, and the higher factorial moments are zero. Further, $u_{ij'}$ is independently distributed of all u's and v's except $u_{i,j'-1}$, $u_{i,j'+1}$, $v_{i-1,j'}$, $v_{i+1,j'}$, $v_{i-1,j'+1}$, $v_{i+1,j'+1}$, and $v_{i'j}$ is independently distributed of all u's and v's excepting two vertically adjacent v's and four horizontally adjacent u's. If

$$s = \Sigma u_{ij'} + \Sigma v_{i'j},$$

then

$$E(s) = \sum_{i,j'} p^2 + \sum_{i',j} p^2$$
$$= (2mn - m - n)\, p^2$$

and $E(s^{(2)}) = 2E$ (the number of ways of selecting any two of the ones included in $\Sigma u_{ij'} + \Sigma v_{i'j}$)

$$= 2\, E\, (uu + uv + vv)$$

involves only the cross products since $E(u^{(2)}) = 0 = E(v^{(2)})$. For products of dependent pairs the expectation is p^3, while for independent pairs it is p^4. Hence one merely needs to count the number of dependent and independent products. Similarly for the third factorial moment one needs consider only products of three first powers of the variates (with expectation p^3), those with two dependent and one independent variates (with expectation p^5), and those with three dependent variates (with expectation p^4).

Thus the second factorial moment can be obtained by counting the number of ways of obtaining two black-black joins from (i) three adjacent points and (ii) two pairs of adjacent points. They are explained below diagrammatically for a 5 x 4 lattice.

```
(1) .   .   .   .   .
    .   X — X — X   .
    .   .   .   .   .
    .   .   .   .   .

(2) .   .   .   .   .
    .   X — X   .   .
            |
    .   .   X   .   .
    .   .   .   .   .

(3) .   .   .   .   .            .   .   .   .   .
    .   X — X   .   .            .   X — X   .   .
    .   .   .   X   .     or     .   .   .   X — X
                |
    .   .   .   X   .            .   .   .   .   .
```

'X' denotes a black point.

'.' denotes any point other than black. The expectations for items (1), (2) and (3) indicated above are

$$(2.1.1)\quad \begin{array}{l} [(m-2)n + (n-2)m]p^3, \\ 4(m-1)\,(n-1)p^3, \\ \tfrac{1}{2}\,[4m^2n^2 - 4mn(m+n) + m^2 + n^2 - 12mn + 13(m+n) - 8]\,p^4, \end{array}$$

respectively Thus

$$\mu'_{[2]} = 2[6mn - 6(m + n) + 4]p^3 + [4m^2n^2 - 4mn(m + n) + m^2n^2 - 12mn + 13(m + n) - 8]p^4. \tag{2.1.2}$$

It can now be seen that

$$\mu'_1 = (2mn - m - n)p^2, \tag{2.1.3}$$

[2]
$$\mu_2 = (2mn - m - n)p^2 + 2(6mn - 6m - 6n + 4)p^3 - (14mn - 13m - 13n + 8)p^4. \tag{2.1.4}$$

Putting $m + n = a$, and $mn = b$, the above expressions reduce to

$$\mu'_1 = (2b - a)p^2, \tag{2.1.5}$$

$$\mu_2 = (2b - a)p^2 + 2(6b - 6a + 4)p^3 - (14b - 13a + 8)p^4. \tag{2.1.6}$$

These substitutions have been continued throughout this Section.

(b) *Non-free sampling* The chances of obtaining r black points in free and non-free sampling are p^r and $n_1^{(r)}/b^{(r)}$ respectively. Therefore it is obvious that the rth factorial moment about zero for non-free sampling distribution of black-black joins can be reduced by substituting $n_1^{(r)}/b^{(r)}$ for p^r in $\mu'_{[r]}$ for free sampling. This substitution gives

$$\mu'_{1(n_1, n_2)} = \frac{(2b - a)\, n_1^{(2)}}{b^{(2)}}, \tag{2.1.7}$$

$$\mu_{2(n_1, n_2)} = \frac{(2b - a)n_1^{(2)}}{b^{(2)}} + \frac{2(6b - 6a + 4)n_1^{(3)}}{b^{(3)}} - \frac{\{(14b - 13a + 8) - (2b - a)^2\}n_1^{(4)}}{b^{(4)}} - \left\{\frac{(2b - a)n_1^{(2)}}{b^{(2)}}\right\}^2, \tag{2.1.8}$$

where $\mu_{r(n_1, n_2)}$ represents the rth moment with n_1 black and n_2 white points on the lattice.

2.2. *Cumulants for the distribution of black-white joins for two colors.* For m points on a line, the author [3] has shown that the first four cumulants of the free and non-free sampling distribution of black-white joins can be obtained from the non-free distributions for $(1, m - 1)$, $(2, m - 2)$, $(3, m - 3)$ and $(4, m - 4)$ black and white points distributed at random. This method is applicable for two and three dimensional lattices also. This can be established from the following considerations.

(i) The rth moment about zero for the free sampling distribution is

$$\sum_0^b p^s q^{b-s} \sum x^r f_x,$$

[2] This result differs slightly from that given by Moran. The correct result is the one given here.

where $b = mn$ and $\Sigma x^r f_x$ is the rth moment for the non-free distribution with s black and $(b - s)$ white points

(ii) $\Sigma x^r f_x$ is the same for the two distributions arising from (1) s black and $(b - s)$ white points and (2) $(b - s)$ black and s white points.

(iii) The rth moment is a polynomial in pq of degree r. This can be seen from the fact that the factorial moment is the sum of the expectations of the different ways of obtaining r black-white joins. The probability of r independent black-white joins is $(2pq)^r$ and this is the highest power of pq

In view of the above conditions, (i) reduces to

$$A'_{1r}pq(p+q)^{(b-2)} + A'_{2r}p^2q^2(p+q)^{b-4} \cdots + A'_{rr}p^rq^r(p+q)^{b-2r} = A'_{1r}pq + A'_{2r}p^2q^2 \cdots + A'_{rr}p^rq^r, \tag{2.2.1}$$

where A'_{1r}, A'_{2r} etc. are determined from the following relations:—

$$\left\{\begin{aligned} S_{r(1,b-1)} &= A'_{1r}, \\ S_{r(2,b-2)} &= A'_{2r} + \binom{b-2}{1} A'_{1r}, \\ S_{r(3,b-3)} &= A'_{3r} + \binom{b-4}{1} A'_{2r} + \binom{b-4}{2} A'_{1r}, \\ S_{r(4,b-4)} &= A'_{4r} + \binom{b-6}{1} A'_{3r} + \binom{b-6}{2} A'_{2r} + \binom{b-6}{3} A'_{1r}, \end{aligned}\right. \tag{2.2.2}$$

where $S_{r(t,b-t)}$ is the rth moment about zero for the non-free distribution with t black and $(b - t)$ white points. This is obvious by comparing the coefficients of p^tq^{b-t} in (i) with (2.2.1).

Therefore the first four cumulants can be calculated by finding the frequency distributions of black-white joins for $(1, b-1)$, $(2, b-2)$, $(3, b-3)$ and $(4, b-4)$ black and white points. These distributions were determined by a systematic examination of the number of black-white joins in all the possible arrangements for the given number of black and white points. The moments of these distributions enable us to determine the A's.

The equations in (2.2.2) give

$$\begin{aligned} A'_{11} &= 2(2b - a), \\ A'_{12} &= 2(8b - 7a + 4), \\ A'_{13} &= 2(32b - 37a + 36), \\ A'_{14} &= 2(128b - 175a + 220), \\ A'_{22} &= 4(a^2 - 4ab + 4b^2 + 13a - 14b - 8), \\ A'_{23} &= 4(21a^2 - 66ab + 48b^2 + 210a - 156b - 228), \\ A'_{33} &= 8(-a^3 + 6a^2b - 12ab^2 + 8b^3 - 39a^2 + 120ab - 84b^2 - 272a + 184b \\ &\quad + 312), \\ A'_{24} &= 4(295a^2 - 760ab + 448b^2 + 2305a - 1304b - 3428), \end{aligned}$$

$$A'_{34} = 8(-42a^3 + 216a^2b - 360ab^2 + 192b^3 - 1410a^2 + 3612ab - 2016b^2 - 7884a + 3648b + 12720),$$

$$A'_{44} = 16(a^4 - 8a^3b + 24a^2b^2 - 32ab^3 + 16b^4 + 78a^3 - 396a^2b + 648ab^2 - 336b^3 + 1643a^2 - 4196ab + 2252b^2 + 7926a - 3084b - 13464),$$

where $a = m + n$, and $b = mn$.

The above values of A''s give the first four moments for free sampling about zero. The cumulants reduce to the following expressions:

$$(2.2.3) \quad \kappa_1 = 2(2b - a)pq,$$

$$(2.2.4) \quad \kappa_2 = 2(8b - 7a + 4)pq - 4(14b - 13a + 8)p^2q^2,$$

$$(2.2.5) \quad \kappa_3 = 2(32b - 37a + 36)pq - 8(90b - 111a + 114)p^2q^2 + 64(29b - 37a + 39)p^3q^3,$$

$$(2.2.6) \quad \kappa_4 = 2(128b - 175a + 220)pq - 4(1784b - 2617a + 3476)p^2q^2 + 32(1548b - 2361a + 3228)p^3q^3 - 32(3126b - 1899a + 6828)p^4q^4.$$

As indicated for black-black joins, the first and the second moments for non-free sampling can be calculated by substituting

$$p^rq^s = n_1^{(r)}n_2^{(s)}/b^{(r+s)}$$

in the uncorrected moments about the origin for free sampling. This is true for all the distributions considered in this paper

Before proceeding to discuss the limiting form of the distribution, it may be noted that the first four cumulants for the free-sampling distribution of black-white joins are linear expressions in a and b. This result is similar to what has been established for the linear lattice (Krishna Iyer, [3]). When the points lie on a line, all the cumulants of the distribution of the number of joins (black-black or black-white) are linear in m (the number of points on the line). This suggests that the higher order cumulants for the distribution of joins in a rectangular lattice also will be linear in a and b, i e. the rth cumulant will be of the form

$$\sum_{s=1}^{r} (L_{rs}b + M_{rs}a + N_{rs})p^sq^s,$$

where L, M and N are independent of a and b. It has not been possible to obtain a formal proof for this statement.

The limiting form of the distribution of the number of black-white joins is now examined on the basis of the cumulants given above Since κ_2, κ_3 and κ_4 are linear in a and b, γ_1 and γ_2 tend to the limit zero as m and n tend to infinity That the higher order γ's also tend to the limit zero can be seen from the fact

that all the cumulants will be linear functions in a and b. Hence the distribution of

$$y = \frac{x - 2(2b - a)pq}{\sqrt{2(8b - 7a + 4)pq - 4(14b - 13a + 8)p^2q^2}}$$

tends to the normal form as m and n tend to infinity, where x is the observed number of black-white joins in a given arrangement of the points

When $p = q = \frac{1}{2}$, the first, second and third cumulants are equal to those obtained for a binomial distribution whose 'n' is $(2b - a)$

As in the case of linear lattices, the distribution of the number of black-white joins in an $m \times n$ rectangular lattice for non-free sampling also will tend to the normal form as m and n tend to infinity.

TABLE 1

Distribution of the number of black-white joins for 2×3 lattice

No of B-W joins	No. of black points 0	1	2	3	4	5	6	Total
0	1	—	—	—	—	—	1	2
1	—	—	—	—	—	—	—	
2	—	4	2	—	2	4	—	12
3	—	2	4	6	4	2	—	18
4	—	—	5	8	5	—	—	18
5	—	—	4	4	4	—	—	12
6	—	—	—	—	—	—	—	
7	—	—	—	2	—	—	—	2
								64

$$\kappa_1 = 7/2, \quad \kappa_2 = 7/4, \quad \kappa_3 = 0, \quad \kappa_4 = \frac{17}{8}.$$

In order to have an idea of the nature of the distribution of the number of black-white joins when $p = q$ or otherwise, the complete distributions for the lattices 2×3, 2×4, 3×3 and 3×4 are given in Tables 1, 2, 3, and 4.

The distributions tabulated in Tables 1, 2, 3 and 4 show that the probability of getting 1 and $(2b - a - 1)$ black-white joins is zero, while for 0 and $(2b - a)$ joins it is not so. But this abnormality will not affect the limiting form of the distribution when m and n tend to infinity because the probability for 0 and $(2b - a)$ black-white joins also tends to zero.

2.3 *First and second moments for the distribution of black-white joins for k colors. Free sampling.* Taking p_1 and p_2 as the probabilities that a point in the lattice is black or white, the expected number of black-white joins is

(2.3 1) $$2(2b - a)\, p_1p_2\,.$$

TABLE 2

Distribution of the number of black-white joins for 2 × 4 lattice

No. of B-W joins	No. of black points 0	1	2	3	4	5	6	7	8	Total
0	1	—	—	—	—	—	—	—	1	2
1	—	—	—	—	—	—	—	—	—	—
2	—	4	2	—	2	—	2	4	—	14
3	—	4	4	4	—	4	4	4	—	24
4	—	—	8	12	8	12	8	—	—	48
5	—	—	12	16	24	16	12	—	—	80
6	—	—	2	12	20	12	2	—	—	48
7	—	—	—	8	8	8	—	—	—	24
8	—	—	—	4	6	4	—	—	—	14
9	—	—	—	—	—	—	—	—	—	—
10	—	—	—	—	2	—	—	—	—	2
										256

$$\kappa_1 = 5, \qquad \kappa_2 = 5/2, \qquad \kappa_3 = 0, \qquad \kappa_4 = \frac{13}{4}.$$

TABLE 3

Distribution of the number of black-white joins for 3 × 3 lattice

No. of B-W joins	No. of black points 0	1	2	3	4	5	6	7	8	9	Total
0	1	—	—	—	—	—	—	—	—	1	2
1	—	—	—	—	—	—	—	—	—	—	—
2	—	4	—	—	—	—	—	—	4	—	8
3	—	4	8	4	—	—	4	8	4	—	32
4	—	1	6	4	12	12	4	6	1	—	46
5	—	—	12	24	12	12	24	12	—	—	96
6	—	—	10	26	36	36	26	10	—	—	144
7	—	—	—	12	36	36	12	—	—	—	96
8	—	—	—	10	13	13	10	—	—	—	46
9	—	—	—	4	12	12	4	—	—	—	32
10	—	—	—	—	4	4	—	—	—	—	8
11	—	—	—	—	—	—	—	—	—	—	—
12	—	—	—	—	1	1	—	—	—	—	2
											512

$$\kappa_1 = 6, \qquad \kappa_2 = 3, \qquad \kappa_3 = 0, \qquad \kappa_4 = 4.5.$$

TABLE 4

Distribution of the number of black-white joins for 4 × 3 lattice

No of B-W joins	No. of black points 0	1	2	3	4	5	6	7	8	9	10	11	12	Total
0	1	—	—	—	—	—	—	—	—	—	—	—	1	2
1	—	—	—	—	—	—	—	—	—	—	—	—	—	
2	—	4	—	—	—	—	—	—	—	—	—	4	—	8
3	—	6	8	2	—	—	2	—	—	2	8	6	—	34
4	—	2	8	8	10	4	—	4	10	8	8	2	—	64
5	—	—	22	28	10	18	16	18	10	28	22	—	—	172
6	—	—	22	46	56	42	30	42	56	46	22	—	—	362
7	—	—	6	52	88	88	120	88	88	52	6	—	—	588
8	—	—	—	50	119	162	156	162	119	50	—	—	—	818
9	—	—	—	28	104	184	186	184	104	28	—	—	—	818
10	—	—	—	6	58	134	192	134	58	6	—	—	—	588
11	—	—	—	—	32	88	122	88	32	—	—	—	—	362
12	—	—	—	—	16	46	48	46	16	—	—	—	—	172
13	—	—	—	—	2	14	32	14	2	—	—	—	—	64
14	—	—	—	—	—	8	18	8	—	—	—	—	—	34
15	—	—	—	—	—	4	—	4	—	—	—	—	—	8
16	—	—	—	—	—	—	—	—	—	—	—	—	—	
17	—	—	—	—	—	—	2	—	—	—	—	—	—	2
														4096

$\kappa_1 = 8.5, \quad \kappa_2 = 4.25, \quad \kappa_3 = 0, \quad \kappa_4 = 0.875.$

TABLE 5

Frequency distribution of the total number of joins between points of different colors for 1 black, 1 white and (mn − 2) red points

No of joins	Frequency
4	28
5	$4(5a - 26)$
6	$2(2a^2 - 25a + 4b + 56)$
7	$2(-4a^2 + 2ab + 17a - 6b - 12)$
8	$4a^2 - 4ab + b^2 - 4a + 3b - 12)$

As in the case of black-black joins, the second factorial moment about zero is twice the sum of the expectations of the different ways of forming two black-white joins and can be determined by the method described in section 2.1.

$$\begin{aligned}(2.3.2)\qquad \mu'_{[2]} &= 2(6b - 6a + 4)p_1p_2(p_1 + p_2)\\ &\quad + 4(a^2 - 4ab + 4b^2 + 13a - 14b - 8)p_1^2p_2^2 .\end{aligned}$$

From this, μ_2 works out to be

$$\begin{aligned}(2.3.3)\qquad \mu_2 &= 2(2b - a)p_1p_2 + 2(6b - 6a + 4)p_1p_2(p_1 + p_2)\\ &\quad - 4(14b - 13a + 8)p_1^2p_2^2 .\end{aligned}$$

2.4. *First and second moments for the distribution of the total number of joins between points of different colors for three colors.* The expectation for free sampling is

$$(2.4.1)\qquad \mu'_1 = 2(2b - a)\Sigma p_rp_s .$$

The coefficients of pq and p^2q^2 in the second moment are the same as those for two colors. The coefficient of $p_1p_2p_3$ can be obtained from the frequency distribution of the total number of joins between points of different colors when there are 1 black, 1 white and $(mn - 2)$ red points in the lattice. See Table 5.

Defining $S_{2(1,1,b-2)} = \Sigma x^2f_x$ for the above distribution,

$$S_{2(1,1,b-2)} = 2(4a^2 - 30ab + 32b^2 + 55a - 54b - 32).$$

As in the case of two colors, the second moment about zero for three colors reduces to the form

$$\begin{aligned}A_{21}(p_1 + p_2 + p_3)^{b-2}\ \Sigma p_rp_s + A_{112}(p_1 + p_2 + p_3)^{b-3}\ p_1p_2p_3 +\\ A_{22}(p + p_1 + p_3)^{b-4}\ \Sigma p_r^2p_s^2 = A_{21}\Sigma p_rp_s + A_{112}p_1p_2p_3 + A_{22}\ \Sigma p_r^2p_s^2 ,\end{aligned}$$

since $p_1 + p_2 + p_3 = 1$.

The coefficient of $p_1^{b-2}p_2p_3$ on the left hand side of the above equation is equal to $S_{2(1,1,b-2)}$, i.e. $S_{2(1,1,b-2)}$ = sum of coefficients of $p_1^{b-2}p_2p_3$ in $A_{21}(p_1 + p_2 + p_3)^{b-2}\ \Sigma p_rp_s$ and $A_{112}(p_1 + p_2 + p_3)^{b-3}p_1p_2p_3$. Therefore the coefficient of $p_1p_2p_3$ in μ_2 is $S_{2(1,1,b-2)}$ − coefficient of $p_1^{mn-2}p_2p_3$ in $2(8b - 7a + 4)(p_1 + p_2 + p_3)^{b-2}\Sigma p_rp_s$ − coefficient of $p_1p_2p_3$ in

$$\begin{aligned}4(2b - a)^2\ (\Sigma p_rp_s)^2 &= S_{2(1,1,b-2)} - 2(8b - 7a + 4)\ (2b - 3) - 8(2b - a)^2\\ &= 4(17a - 19b - 10).\end{aligned}$$

It can now be seen that

$$\begin{aligned}(2.4.2)\qquad \mu_2 &= 2(8b - 7a + 4)\Sigma p_rp_s - 4(14b - 13a + 8)\Sigma p_r^2p_s^2\\ &\quad - 4(19b - 17a + 10)p_1p_2p_3 .\end{aligned}$$

2.5 *First and second moments for the distribution of the total number of joins between points of different colors for k colors.* As in the previous cases, the expectation for free sampling is

$$(2.5.1)\qquad 2(2b - a)\Sigma p_rp_s .$$

The coefficients of $\Sigma p_r p_s$, $\Sigma p_r p_s p_t$ and $\Sigma p_r^2 p_s^2$ in the second moment are the same as those for three colors. The coefficient of $\Sigma p_r p_s p_t p_u$ is determined by finding the distribution of joins between points of different colors when there are 1 black, 1 white, 1 red and $mn - 3$ green points in the lattice. See Table 6.

$$S_{2(1,1,1,mn-3)}$$
$$= 2(12a^2b - 69ab^2 + 72b^3 - 36a^2 + 330ab - 342b^2 - 408a + 318b + 240).$$

The coefficient of $\Sigma p_r p_s p_t p_u$ in μ_2 can be obtained on the same lines as explained for three colors and is equal to $S_{2(1,1,1,mn-3)}$ − coefficient of $p_r^{(mn-3)} p_s p_t p_u$ in the homogeneous expression of degree mn in μ_2' for three colors $+ 8(2b - a)^2$

$$= 8(14b - 13a + 8)$$

TABLE 6

Frequency distribution of the total number of joins between points of different colors when there are 1 black, 1 white, 1 red and (mn − 3) green points

No of joins	Frequency
6	240
7	$12(19a - 112)$
8	$12(6a^2 - 78a + 7b + 208)$
9	$4(2a^3 - 57a^2 + 15ab + 310a - 66b - 444)$
10	$6(-4a^3 + 2a^2b + 36a^2 - 21ab + 2b^2 - 86a + 36b + 72)$
11	$6(4a^3 - 4a^2b + ab^2 - 6a^2 + 8ab - 2b^2 - 10a - 40)$
12	$(-8a^3 + 12a^2b - 6ab^2 + b^3 - 24a^2 + 18ab - 3b^2 + 44a - 34b + 192)$

It follows now that

$$\mu_2 = 2(8b - 7a + 4)\Sigma p_r p_s - 4(19b - 17a + 10)\Sigma p_r p_s p_t$$
$$- 4(14b - 13a + 8)\Sigma p_r^2 p_s^2 + 8(14b - 13a + 8)\Sigma p_r p_s p_t p_u . \tag{2.5.2}$$

In general the cumulants[3] for free sampling involve b and a in the first degree only, and therefore, when m and n are large, the distribution tends to the normal form. If x is the observed total number of joins between points of different colors, the distribution of

$$\frac{x - 2(2b - a)\Sigma p_r p_s}{\sqrt{b}}$$

[3] The author has recently obtained the third and fourth cumulants for this distribution. They are linear functions of the dimensions of the lattice. The results will be published in an early issue of the *Ind J Agric. Stat.*

tends to the normal form with

$$16\Sigma p_r p_s - 76\Sigma p_r p_s p_t - 56\Sigma p_r^2 p_s^2 + 112\Sigma p_r p_s p_t p_u ,$$

as its variance for large values of m and n.

For non-free sampling also, the distribution of

$$\frac{x - 2(2mn - m - n)\Sigma e_r e_s}{\sqrt{mn}},$$

where $e_r = n_r/mn$, approaches the normal form having

$$4\Sigma e_r e_s e_t + 8\Sigma e_r^2 e_s^2 - 16\Sigma e_r e_s e_t e_u$$

as its variance. The error of this variance will be about 5% or less when m and n are greater than 35.

3. Three- and higher-dimensional lattices. This section deals with the first and the second moments for the distribution of black-black, black-white and the total number of joins between points of different colors for three- and higher-dimensional lattices. Besides these, the third and the fourth cumulants for the distribution of black-white joins in a three-dimensional lattice with points of two colors are also given.

3.1 *First and second moments for the distribution of black-black joins. Free sampling.* Let $E_3(1)$ be the expectation of the number of black-black joins for a lattice of sides l, m and n. Further let A_2 and A_3 be the number of ways of obtaining a black-black join in $m \times n$ and $l \times m \times n$ lattices. Then

$$E_3(1) = A_3 p_1^2 ,$$

$$A_3 = A_2 l + mn(l - 1),$$

and

$$A_2 = (2mn - m - n).$$

Therefore

$$E_3(1) = (3lmn - lm - mn - nl)\, p_1^2 . \tag{3.1.1}$$

For the sake of convenience all the results for the three-dimensional lattice are expressed after making the following substitutions:

$$c = l + m + n,$$

$$d = lm + mn + nl,$$

$$e = lmn$$

$E_3(1)$ in terms of c, d and e is

$$(3e - d)p_1^2 .$$

The expectation of the number of black-black joins for a lattice of r dimensions $(l_1 \times l_2 \times \cdots l_r)$ is given by

(3.1.2) $$E_r(1) = (rl_1l_2 \cdots l_r - \Sigma l_1l_2 \cdots l_{r-1})\, p_1^2,$$

where $\Sigma l_1l_2 \cdots l_{(r-1)}$ is the sum of the product of the sides taken $(r - 1)$ at a time.

It has been pointed out before that the second factorial moment is twice the sum of the expectations of the different ways of forming two black-black joins. Using this fact, if $2B_2$, $2B_3$, etc. are the coefficients of p^3 in the second factorial moment for two-, three- and higher-dimensional lattices, it will be found by direct enumeration made in succession from lattices of lower dimensions that

$$B_r = B_{(r-1)}l_r + 4A_{(r-1)}\,(l_r - 1) + l_1l_2 \cdots l_{(r-1)}\,(l_r - 2).$$

This can be established from the following considerations. 1) Two black-black joins can be obtained from three black points situated close to one another and the chance of having three black points in a specified manner is p^3. 2) The number of ways of getting two black-black joins from three points in the lattice is

$$B_{(r-1)}l_r + 4A_{(r-1)}(l_r - 1) + l_1l_2 \cdots l_{(r-1)}\,(l_r - 2).$$

C_r, the coefficient of p^4 in the corrected second moment, is given by the equation

$$C_r = -(2B_r + A_r).$$

This follows from the fact that the sum of the coefficients of p^3 and p^4 in the uncorrected factorial moment, about zero, is twice the number of ways of selecting two joins from the total number of joins in the lattice which is $(A_r - 1)$. Thus

(3.1.3) $$A_rp_1^2 + 2B_rp_1^3 + C_rp_1^4$$

is the corrected second moment for the distribution of black-black joins in a lattice of r dimensions. For an $l \times m \times n$ lattice

(3.1.4) $$\mu_2 = (3e - d)\,p_1^2 + 2(15e - 10d + 4c)\,p_1^3 - (33e - 21d + 8c)\,p_1^4.$$

3.2. *Cumulants for the distribution of black-white joins for two colors.* The first four cumulants for free and non-free sampling distributions in an $l \times m \times n$ lattice can be determined from the frequency distributions of black-white joins for $(1, lmn - 1)$, $(2, lmn - 2)$, $(3, lmn - 3)$ and $(4, lmn - 4)$ black and white points by the method described for linear rectangular lattices. If

$$\mu_r' = A_{1r}''pq + A_{2r}''p^2q^2 + \cdots + A_{rr}''p^rq^r,$$

the first three distributions give the coefficients of pq, p^2q^2 and p^3q^3 in the first three moments about zero. The three cumulants calculated from these moments are given below in terms of c, d, and e for free sampling.

(3.2.1) $$\kappa_1 = 2(3e - d)pq,$$

(3.2.2) $$\kappa_2 = 2(18e - 11d + 4c)pq - 4(33e - 21d + 8c)p^2q^2,$$

(3.2.3) $$\begin{aligned}\kappa_3 = {} & 2(108e - 91d + 60c - 24)pq \\ & + 8(327e - 288d + 198c - 84)p^2q^2 \\ & + 32(219e - 197d + 138c - 60)p^3q^3.\end{aligned}$$

The calculation of the fourth cumulant by the direct method of finding the frequency distribution of the number of black-white joins for 4 black and $(lmn-4)$ white points was found to be very laborious and therefore this has been calculated by a special method. The coefficients of pq, p^2q^2 and p^3q^3 have been determined, as in other cases, by finding $\Sigma x^4 f_x$ for the first three distributions. These coefficients reduce to a linear form in c, d and e. Now the fourth cumulant, being a linear function of these quantities, the coefficient of p^4q^4 involves c, d and e in the first degree only and therefore this can be assumed to be of the form

$$\alpha e + \beta d + \gamma c + \delta,$$

where α, β, γ and δ are constants. No simple proof can be given here regarding the linear assumption of the cumulants. It may be observed that this is true of the first four cumulants for linear and rectangular lattices. The author [3] has already provided a general proof of this assumption for the linear lattice and he hopes to extend this for the higher dimensional lattices in the near future.[4]

The constants α, β, γ, and δ can be determined by finding κ_4 for $p = q = \frac{1}{2}$ from the frequency distributions of black-white joins for $2 \times 2 \times 2$, and $2 \times 3 \times 3$ lattices for two colors as given in Tables 7 and 8.

When $p = q = \frac{1}{2}$, κ_4 reduces to the form $a'e + b'd + c'c + d'$, where a', b', c' and d' are constants. In view of this relation, if m and n are fixed, and l takes values 1, 2, 3, etc., the values of κ_4 for the different lattices should be in arithmetic progression. This can be seen by comparing the values of κ_4 for the lattices $1 \times 2 \times 2$, $2 \times 2 \times 2$ and $3 \times 2 \times 2$ which are 1, 7.5 and 14, respectively. Using this property, it is possible to find κ_4 for a lattice of any size from the complete distribution of the lattices $1 \times 2 \times 2$, $1 \times 2 \times 3$, $1 \times 3 \times 3$, and $2 \times 2 \times 2$ given before. Thus κ_4 for $2 \times 2 \times 2$, $2 \times 2 \times 3$, $3 \times 3 \times 2$ and $3 \times 3 \times 3$ lattices are 7.5, 14, 25.875 and 47.25 respectively. Now α, β, γ and δ can be obtained by equating the general expression for the fourth cumulant to the values given above for the corresponding values of l, m and n and putting $p = q = \frac{1}{2}$. The equations giving the values of α, β, γ and δ are

(3.2.4) $$\begin{cases} 8\theta_1 + 12\theta_2 + 6\theta_3 + \theta_4 = 7.5, \\ 12\theta_1 + 16\theta_2 + 7\theta_3 + \theta_4 = 14.0, \\ 18\theta_1 + 21\theta_2 + 8\theta_3 + \theta_4 = 25.875, \\ 27\theta_1 + 27\theta_2 + 9\theta_3 + \theta_4 = 47.25, \end{cases}$$

[4] This proof has been obtained recently and will be published soon.

where $\theta_1 = \dfrac{32 \times 19176 + \alpha}{256}$, $\theta_2 = \dfrac{-32 \times 21638 + \beta}{256}$,

$\theta_3 = \dfrac{32 \times 20952 + \gamma}{256}$, and $\theta_4 = \dfrac{-32 \times 16128 + \delta}{256}$.

They give

$$\alpha = -32 \times 19143, \qquad \beta = 32 \times 21615,$$

$$\gamma = -32 \times 20940, \text{ and } \delta = 32 \times 16128.$$

TABLE 7

Frequency distribution of black-white joins, 2 × 2 × 2 lattice for two colors

No. of black-white joins	No. of black points									Total
	0	1	2	3	4	5	6	7	8	
0	1	—	—	—	—	—	—	—	1	2
1	—	—	—	—	—	—	—	—	—	—
2	—	—	—	—	—	—	—	—	—	—
3	—	8	—	—	—	—	—	8	—	16
4	—	—	12	—	6	—	12	—	—	30
5	—	—	—	24	—	24	—	—	—	48
6	—	—	16	—	32	—	16	—	—	64
7	—	—	—	24	—	24	—	—	—	48
8	—	—	—	—	30	—	—	—	—	30
9	—	—	—	8	—	8	—	—	—	16
10	—	—	—	—	—	—	—	—	—	—
11	—	—	—	—	—	—	—	—	—	—
12	—	—	—	—	—	—	—	—	—	2
										256

$\kappa_1 = 6, \quad \kappa_2 = 3, \quad \kappa_3 = 0, \quad \kappa_4 = 7.5$

Thus the general formula for the fourth cumulant is

$$(3.2.5) \quad \begin{cases} \kappa_4 = 2(648e - 671d + 604c - 432)pq \\ \quad -4(9996e - 10857d + 10196c - 7632)p^2 q^2 \\ \quad +32(9144e - 10167d + 9732c - 7416)p^3 q^3 \\ \quad -32(19143e - 21615d + 20940c - 16128)p^4 q^4. \end{cases}$$

For a lattice of sides $l_1, l_2, \cdots l_r$ in r dimensions, the first two moments for the distribution of black-white joins for free sampling are as follows:

(3.2.6) $$\mu_1' = 2A_r pq,$$

(3.2.7) $$\mu_2 = 2(A_r + B_r)pq + 4C_r p^2 q^2.$$

Like the distributions for linear and rectangular lattices, when l, m and n tend to infinity, γ_1 and γ_2 will tend to zero and therefore the distribution of black-white joins for an $l \times m \times n$ lattice also tends to the normal form. The remarks made in connection with the distribution of black-white joins for a rectangular lattice are true here also Here the frequencies for 1, 2, $[(3e - d) - 2]$ and $[(3e - d) - 1]$ black-white joins are zero, while for 0 and $(3e - d)$

TABLE 8

Frequency distribution of black-white joins for 2 × 3 × 3 lattice for two colors

No. of black-white joins	No. of black points													Total
	0	1	2	3	4	5	6	7	8	9	10	11	12	
0	1	—	—	—	—	—	—	—	—	—	—	—	1	2
1	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3	—	8	—	—	—	—	—	—	—	—	—	8	—	16
4	—	4	8	—	2	—	—	—	2	—	8	4	—	28
5	—	—	8	8	—	—	—	—	—	8	8	—	—	32
6	—	—	24	20	8	8	12	8	8	20	24	—	—	132
7	—	—	24	48	40	40	16	40	40	48	24	—	—	320
8	—	—	2	52	81	56	68	56	81	52	2	—	—	450
9	—	—	—	40	104	112	144	112	104	40	—	—	—	656
10	—	—	—	44	100	188	160	188	100	44	—	—	—	824
11	—	—	—	8	88	144	176	144	88	8	—	—	—	656
12	—	—	—	—	36	108	162	108	36	—	—	—	—	450
13	—	—	—	—	24	88	96	88	24	—	—	—	—	320
14	—	—	—	—	12	28	52	28	12	—	—	—	—	132
15	—	—	—	—	—	8	16	8	—	—	—	—	—	32
16	—	—	—	—	—	4	20	4	—	—	—	—	—	28
17	—	—	—	—	—	8	—	8	—	—	—	—	—	16
18	—	—	—	—	—	—	—	—	—	—	—	—	—	—
19	—	—	—	—	—	—	—	—	—	—	—	—	—	—
20	—	—	—	—	—	—	2	—	—	—	—	—	—	2
														4096

$\kappa_1 = 10, \qquad \kappa_2 = 5, \qquad \kappa_3 = 0, \qquad \kappa_4 = 14$

they are two But this irregularity will not affect the limiting form of the distribution since the relative frequencies tend to zero.

3 3. *First and second moments for the distribution of black-white joins for k colors in an r-dimensional lattice.* The results for free sampling follow easily from a consideration of the expectations of the various ways of obtaining one and two black-white joins. The expectation of the number of black-white joins is

(3 3.1) $$2\, A_r p_1 p_2\,.$$

The expectation for two black-white joins is

$$B_r p_1 p_2 (p_1 + p_2) + 4\left\{\frac{A_r(A_r - 1)}{2} - B_r\right\} p_1^2 p_2^2.$$

From this it will follow that the second moment

(3.3.2) $$\mu_2 = 2A_r p_1 p_2 + 2B_r p_1 p_2 (p_1 + p_2) + 4C_r p_1^2 p_2^2$$

3.4. *First and second moments for the distribution of the total number of joins between points of different colors for an $l \times m \times n$ lattice for three colors.* The expectation for free sampling is

(3.4.1) $$2(3e - d)\Sigma p_r p_s$$

TABLE 9

Distribution of joins between points of different colors for 1 black, 1 white and $(lmn - 2)$ red points

No. of joins	Frequency for lattices			
	$2 \times 2 \times 2$	$2 \times 2 \times 3$	$2 \times 3 \times 3$	$3 \times 3 \times 3$
5	24	16	8	-
6	32	56	80	104
7	—	56	104	144
8	—	4	96	276
9	—	—	18	112
10	—	—	—	66
Total	56	132	306	702
$\Sigma x^2 f_x$ about zero	1752	5416	15778	41136

The second moment will involve terms in $\Sigma p_r p_s$, $p_1 p_2 p_3$ and $\Sigma p_r^2 p_s^2$. The coefficients of $\Sigma p_r p_s$ and $\Sigma p_r^2 p_s^2$ are the same as those for two colors. The coefficient of $p_1 p_2 p_3$ can be determined by finding the frequency distribution of joins between points of different colors when the lattice consists of 1 black, 1 white and $(lmn - 2)$ red points. But this straightforward method is cumbersome and hence the coefficient of $p_1 p_2 p_3$ has been determined by finding the distribution for the special lattices $2 \times 2 \times 2$, $2 \times 2 \times 3$, $2 \times 3 \times 3$, and $3 \times 3 \times 3$. These results are shown in Table 9.

The coefficients of $p_1 p_2 p_3$ in the corrected second moment for the above lattices are obtained by subtracting $2(18e - 11d + 4c)(2e - 3) + 8(3e - d)^2$ from the moments noted above. This can be seen to be so by comparing the above expression with the quantity subtracted from the uncorrected second moment for a two dimensional lattice in section 2.4. The coefficients so obtained for $2 \times 2 \times 2$,

$2 \times 2 \times 3$, $2 \times 3 \times 3$, and $3 \times 3 \times 3$ lattices are -336, -640, -1184 and -2142 respectively. Now the coefficient of $p_1p_2p_3$ in the corrected second moment is of the form

$$\alpha' e + \beta' d + \gamma' c + \delta'.$$

The equations obtained by equating this expression to -336, -640, -1184 and -2142 for the respective lattices give $\alpha' = -174$, $\beta' = 108$, $\gamma' = -40$

TABLE 10

Distribution of joins between points of different colors when there are 1 black, 1 white, 1 red and (lmn-3) green points

No of joins	Frequency for lattices			
	$2 \times 2 \times 2$	$2 \times 2 \times 3$	$2 \times 3 \times 3$	$3 \times 3 \times 3$
7	144	48	—	—
8	144	312	288	72
9	48	480	912	1344
10	—	432	1344	2664
11	—	48	1560	4392
12	—	—	720	4584
13	—	—	72	3168
14	—	—	—	1206
15	—	—	—	120
Total.	336	1320	4896	17550
$\Sigma x^2 f_x$ about zero . .	20160	110208	531312	2370168

and $\delta' = 0$. Thus the second moment for a lattice with points in three colors is

$$\begin{aligned} &2(18e - 11d + 4c)\Sigma p_r p_s \\ (3.4.2) \qquad &-2(87e - 54d + 20c)p_1p_2p_3 \\ &-4(33e - 21d + 8c)\Sigma p_r^2 p_s^2 . \end{aligned}$$

3.5. *First and second moments for the distribution of the total number of joins between points of different colors in an $l \times m \times n$ lattice for four or more colors.* The expectations for free sampling are given by the same expression as for three colors. The coefficients of $\Sigma p_r p_s$, $\Sigma p_r p_s p_t$ and $\Sigma p_r^2 p_s^2$ in the corrected second moment are also the same as in section 3.4. The coefficient of $\Sigma p_r p_s p_t p_u$ can be determined by the method described in section 3.4 for $\Sigma p_1 p_2 p_3$ from the frequency distributions of joins (Table 9) between points of different colors for $2 \times 2 \times 2$, $2 \times 2 \times 3$, $2 \times 3 \times 3$ and $3 \times 3 \times 3$ lattices when they consist of 1 black, 1 white, 1 red and $(e - 3)$ green points.

The coefficient of $\Sigma p_r p_s p_t p_u$ in the corrected second moment is obtained by subtracting (obtained in the same way as for the two dimensional lattice in section 2.5)

$$\begin{aligned}&6(18c - 11d + 4e)(e - 2)^2\\&+(3c - 8)[2(-87c + 54d - 20e) + 8(3c - d)^2]\\&-8(3c - d)^2\end{aligned}$$

from the uncorrected values. The values so obtained for the four lattices are $480(2\times 2 \times 2)$, $928(2 \times 2 \times 3)$, $1736(2 \times 3 \times 3)$ and $3168(3 \times 3 \times 3)$. The coefficient of $p_r p_s p_t p_u$, as in other cases, being of the form

$$\alpha'' c + \beta'' d + \gamma'' e + \delta'',$$

α'', β'', γ'' and δ'' can be determined by equating the above expression to 480, 928, 1736 and 3168 for the respective lattices. The coefficient so obtained is

$$8(33c - 21d + 8e).$$

Hence the second moment for free sampling when the lattice contains points of four or more colors is

$$\text{3.5.1)}\quad \begin{aligned}&2(18c - 11d + 4e)\Sigma p_r p_s\\&-2(87c - 54d + 20e)\Sigma p_r p_s p_t\\&-4(33c - 21d + 8e)\Sigma p_r^2 p_s^2\\&+8(33c - 21d + 8e)\Sigma p_r p_s p_t p_u\,.\end{aligned}$$

In general, it will be found that the cumulants involve terms in c, d, e and an absolute term only. Therefore when l, m and n tend to infinity and $p_1, p_2, p_3 \cdot$ are finite, the distribution of $R - 2(3e - d)\Sigma p_r p_s$, where R is the total number of joins of points of different colors, tends to the normal form. When l, m and n are large,

$$\frac{R - 2(3e - d)\Sigma p_r p_s}{\sqrt{e}}$$

can be considered to be normally distributed with

$$\text{(3 5.2)}\qquad 36\Sigma p_r p_s - 174\Sigma p_r p_s p_t - 132\Sigma p_r^2 p_s^2 + 264\Sigma p_r p_s p_t p_u$$

as its variance.

The distribution for non-free sampling here also tends to the normal form for the same reasons given for the rectangular lattice. As in free sampling, for large values of l, m and n

$$\frac{R - 2(3e - d)\Sigma e_r e_s}{\sqrt{e}}$$

is distributed normally with the variance

$$\text{(3.5 4)}\qquad 6\Sigma e_r e_s e_t + 12\Sigma e_r^2 e_s^2 - 24\Sigma e_r e_s e_t e_u\,,$$

where R is the observed number of joins for a given distribution of the points and $c_r = \frac{n_r}{lmn}$. The error in this variance will be about 5% or less when l, m and n are greater than 36.

We may conclude this section by giving the first and the second moments for free sampling with k colors for an r-dimensional lattice.

$$\mu_1' = 2A_r \Sigma p_r p_s, \tag{3.5.5}$$

$$\begin{aligned} \mu_2 &= 2(A_r + B_r)\Sigma p_r p_s \\ &+ 2(3B_r + 4C_r)\Sigma p_r p_s p_t \\ &+ 4C_r \Sigma p_r^2 p_s^2 - 8C_r \Sigma p_r p_s p_t p_u , \end{aligned} \tag{3.5.6}$$

where A_r, B_r and C_r are as defined in section 3.1

This can be seen from the following facts:

(1) The coefficients of $\Sigma p_r p_s$ and $\Sigma p_r^2 p_s^2$ are the same as for two colors

(2) The coefficient of $\Sigma p_r p_s p_t$ is the number of ways of getting two joins of different colors from combination of points not included in $\Sigma p_r p_s p_t p_u$. This can be had from three points of three different colors close together and four points of three different colors separated into groups of two each such that each group will give one join. The number of arrangements of the first kind is $3!B_r$. For the second kind it is $8(A_r^2 + C_r)$. Subtracting from the total number, the contribution of $\Sigma p_r p_s p_t$ in the correction factor $4A_r^2(\Sigma p_r p_s)^2$, the coefficient of $\Sigma p_r p_s p_t$ in the second moment works out to be

$$2(3B_r + 4C_r).$$

(3) The coefficient of $\Sigma p_r p_s p_t p_u$, as in all other cases dealt before, is twice that of $\Sigma p_r^2 p_s^2$ with an opposite sign.

Acknowledgements. The author's thanks are due to Dr. D. J. Finney for suggesting this problem and for all the facilities and help given to him in carrying out the investigations discussed in this paper. The author is also grateful to Mr. P A P. Moran for explaining the results of his investigations on this problem and for the interest taken by him in the course of the research.

REFERENCES

[1] V. Bortkiewicz, *Die Iteration*, Berlin, 1917.

[2] R. C. Bose, "The patch number problem," *Sci Cult*, Vol. 12 (1946), p. 199.

[3] P. V Krishna Iyer, "The theory of probability distribution of points on a line," *Indian J.Agric. Stat.*, Vol. 1 (1948).

[4] H. Levene, "A test of randomness in two dimensions," *Ann. Math. Stat.*, Vol. 17 (1946), p. 500

[5] P A Moran, "Random associations on a lattice," *Proc. Cam. Phil. Soc*, Vol 43 (1947), p. 321.

[6] P A. Moran, "The interpretation of statistical maps," *J. Roy Stat. Soc*, B, Vol 10 (1948), p 243

MINIMAX ESTIMATES OF THE MEAN OF A NORMAL DISTRIBUTION WITH KNOWN VARIANCE

By J. Wolfowitz[1]

Columbia University

Summary. It is proved that the classical estimation procedures for the mean of a normal distribution with known variance are minimax solutions of properly formulated problems. A result of Stein and Wald [1] is an immediate consequence. Other such optimum properties follow. Sequential and non sequential problems can be treated in this manner. Interval and point estimation are discussed.

1. Sequential estimation by an interval of given length l. In this section we shall consider the problem of sequentially estimating the mean of a normal distribution with known variance by an interval of fixed length l. Without loss of generality we shall take the known variance to be unity. Such a sequential estimation procedure, which we shall designate generically by G, is a rule which says a) when to terminate taking random, independent observations on the normal chance variable with unknown mean $\xi(-\infty < \xi < \infty)$ and variance 1, and when this termination is to occur after the observations $x_1, \cdots, x_n$ have been obtained, gives b) the center of the estimating interval of length l as a function of $x_1, \cdots, x_n$ Let $\alpha(\xi, G)$ be the probability under G that the estimating interval will contain ξ, and let $n(\xi, G)$ be the expected number of observations when ξ is the mean and G is the estimation procedure (It is assumed that G is such that $\alpha(\xi, G)$ and $n(\xi, G)$ exist for all ξ).

Define

$$q(\xi, G) = 1 - \alpha(\xi, G),$$

and for fixed $c > 0$

$$W(\xi, G) = q(\xi, G) + cn(\xi, G). \tag{1.1}$$

Let $C(N, l)$ ($l > 0$, N a positive integer) be the classical non-sequential estimation procedure where one takes the fixed number N of observations, and estimates the mean by the interval $\left(\bar{x} - \frac{l}{2}, \bar{x} + \frac{l}{2}\right)$, where $\bar{x}$ is the sample mean. For p such that $0 < p \leq 1$, let $C(p, N, l)$ be the following estimation procedure: A chance experiment with two outcomes, N and $N + 1$, of respective probabilities p and $1 - p$, is performed. One then proceeds according to $C(i, l)$, where $i(= N, N + 1)$ is the outcome of the experiment. Finally define

$$M(y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty e^{-\frac{1}{2}z^2}\, dz.$$

[1] Research under a contract with the Office of Naval Research

Let us assume for a moment that the unknown ξ is itself a chance variable, normally distributed with mean zero and variance σ^2, and let us obtain a procedure G which minimizes

$$(1.2)\quad E\{q(\xi, G) + c\,n(\xi, G)\} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\{q(y, G) + cn(y, G)\}\exp\left[\frac{-y^2}{2\sigma^2}\right]dy$$

Let $x_1, \cdots, x_m$ be m independent observations on a normal chance variable with mean ξ and variance 1. Let

$$\bar{x} = \frac{\sum_1^m x_i}{m}$$

The a posteriori distribution of ξ, given $x_1, \cdots, x_m$, is easily verified (or see [1], eqs. (19) and (20)) to be normal with mean

$$(1.3)\qquad \bar{x}\left[1 + \frac{1}{m\sigma^2}\right]^{-1}$$

and variance

$$(1.4)\qquad \left[m + \frac{1}{\sigma^2}\right]^{-1}.$$

Thus if we stop after m observations the best procedure from the point of view of minimizing (1.2) is to put the center of the estimating interval of length l at the point (1.3). The conditional expected value of $q(\xi)$ is then

$$(1.5)\qquad Q(x_1, \cdots, x_m \mid \sigma^2) = 2M\left(\frac{l}{2}\sqrt{m + \frac{1}{\sigma^2}}\right).$$

Thus $Q(x_1, \cdots, x_m)$ is a function only of m and σ^2. Define

$$(1.6)\qquad R(m, \sigma^2) = 2M\left(\frac{l}{2}\sqrt{m + \frac{1}{\sigma^2}}\right) - 2M\left(\frac{l}{2}\sqrt{m + 1 + \frac{1}{\sigma^2}}\right).$$

We note that $R(m, \sigma^2)$ is, for fixed σ, a decreasing function of m. We conclude that a best decision as to whether or not to take another observation must be based on the value of $R(m, \sigma^2)$. If $R(m, \sigma^2) > c$ take another observation, if $R(m, \sigma^2) < c$ do not take another observation, if $R(m, \sigma^2) = c$ take either action at pleasure. Hence, if c is such that $R(N, \sigma^2) \le c \le R(N - 1, \sigma^2)$, a best procedure from the point of view of minimizing (1.2) is to take exactly N observations. This integer N is a function of c and σ^2, thus: $N(c, \sigma^2)$. In the next paragraph we shall show that $N(c, \sigma^2)$ can be defined for every positive c and σ^2. It is clearly a function which takes at most two values. We shall denote by $G(\sigma^2)$ the estimation procedure described above which minimizes (1.2). It consists of taking the fixed number $N(c, \sigma^2)$ of observations and putting the center of the estimating interval of length l at the point (1.3). Where $N(c, \sigma^2)$ is double-valued we may take either value at pleasure. We verify that the value of (1.2) is the same for either choice.

We now verify that $N(c, \sigma^2)$ can be defined for all positive c and σ^2. We have remarked earlier that $R(m, \sigma^2)$ is, for fixed σ^2, a monotonically decreasing function of m. We note that

$$\lim_{m\to\infty} R(m, \sigma^2) = 0.$$

When $c > R(0, \sigma^2)$ we take no observations whatever and take $\bar{x} = 0$. When $c = R(0, \sigma^2)$ we take zero or one observation at pleasure.

Without difficulty we compute

$$W(\xi, G(\sigma^2)) = W(\xi, \sigma^2) = cN + M\left(\sqrt{N}\,\frac{l}{2}\left[1 + \frac{1}{N\sigma^2}\right] - \frac{\xi}{\sqrt{N}\,\sigma^2}\right) + M\left(\sqrt{N}\,\frac{l}{2}\left[1 + \frac{1}{N\sigma^2}\right] + \frac{\xi}{\sqrt{N}\,\sigma^2}\right)$$

where for typographical simplicity we have written N for $N(c, \sigma^2)$. For fixed c and σ^2 the minimum of $W(\xi, \sigma^2)$ occurs at $\xi = 0$. Also $W(0, \sigma^2)$ is a monotonically increasing function of σ^2. If $N(c, \infty) > 0$ then, as $\sigma^2 \to \infty$ it approaches the limit

$$cN(c, \infty) + 2M\left(\frac{l}{2}\sqrt{N(c, \infty)}\right),$$

which is the constant value of

$$W(\xi, C(N(c, \infty), l)).$$

We therefore conclude that $C(N(c, \infty), l)$ is a minimax estimating procedure of type G, i.e.,

$$W(\xi, C(N(c, \infty), l)) = \inf_G \sup_\xi W(\xi, G)$$

for any $c > 0$. (The case $N(c, \infty) = 0$ may be verified separately. We define $\bar{x} \equiv 0$ for $C(0, l)$).

Conversely, let N_0 be a given non-negative integer. Then $C(N_0, l)$ is a minimax estimating procedure G for all $W(\xi, G)$ for which c satisfies

$$R(N_0, \infty) \le c \le R(N_0 - 1, \infty).$$

(We define $R(-1, \infty) = \infty$.) Thus we can say: For every $c > 0$ there exists a classical estimation procedure $C(N, l)$ with integral N such that

$$W(\xi, C(N, l)) = \inf_G \sup_\xi W(\xi, G).$$

For every integral N we can find at least one $c > 0$ such that the above equation holds. A method of finding N, given c, and of finding c, given N, has been described above. (We have taken the liberty of calling $C(0, l)$ a classical procedure.

Let α_0 be a given number such that

$$1 - 2M\left(\frac{l}{2}\right) \le \alpha_0 < 1.$$

Define p_0, $0 < p_0 \leq 1$, and a positive integral N_0 uniquely by

$$\alpha_0 = p_0 \left(1 - 2M\left(\sqrt{N_0}\, \frac{l}{2}\right)\right) + (1 - p_0) \left(1 - 2M\left(\sqrt{N_0 + 1}\, \frac{l}{2}\right)\right).$$

Let

$$c_0 = R(N_0, \infty).$$

For $c = c_0$ we verify readily that both $C(N_0, l)$ and $C(N_0 + 1, l)$ are minimax estimating procedures G, so that

$$\begin{aligned} W(\xi, C(N_0, l)) &= W(\xi, C(N_0 + 1, l)) \\ &= p_0 W(\xi, C(N_0, l)) + (1 - p_0) W(\xi, C(N_0 + 1, l)) \\ &= (1 - \alpha_0) + c_0[p_0 N_0 + (1 - p_0)(N_0 + 1)] \\ &= (1 - \alpha_0) + c_0[N_0 + (1 - p_0)]. \end{aligned}$$

Therefore, for any G whatever,

$$\begin{aligned} (1 - \alpha_0) + c_0[N_0 + (1 - p_0)] &\leq \sup_\xi \{q(\xi, G) + c_0\, n(\xi, G)\} \\ &\leq \sup_\xi q(\xi, G) + c_0 \sup_\xi n(\xi, G). \end{aligned}$$

Hence

$$\sup_\xi q(\xi, G) \leq 1 - \alpha_0$$

implies

$$\sup_\xi n(\xi, G) \geq N_0 + (1 - p_0),$$

a result first proved by Stein and Wald [1].
Also

$$\sup_\xi n(\xi, G) \leq N_0 + (1 - p_0)$$

implies

$$\sup_\xi q(\xi, G) \geq 1 - \alpha_0,$$

a result also proved in [1].

2. A sequential upper bound for the mean. The fact that in the last section l was a constant made matters simpler, as we see when we begin to consider the problem of a sequential upper bound for $\xi(-\infty < \xi < \infty)$. This of course means that we wish to use as estimating interval the interval $(-\infty, L(x_1, \cdots, x_n))$ where L is a function of the observations $x_1, \cdots, x_n$, and n (a chance variable) is the number of observations before the process of taking observations is terminated. What is wanted now is a suitable definition of the "length" of this in-

terval. Also we shall admit the possibility that it might be in some sense advantageous to have intervals of varying length; this poses the problem of optimum choice of the function $L(x_1, \cdots, x_n)$.

As before, let ξ be the mean of a normal distribution with unit variance. Let T be the generic estimation procedure which consists of a rule for terminating the taking of observations, and of a function $L_T(x_1, \cdots, x_n)$ which is used to estimate ξ by the interval $(-\infty, L_T)$. Define

$$q(\xi, T) = P\{L_T \leq \xi\},$$
$$\lambda(\xi, T) = E(L_T - \xi)^2,$$

and

$$W(\xi, T) = q(\xi, T) + k\lambda(\xi, T) + cn(\xi, T), \tag{2.1}$$

where c and k are positive constants. (We admit only such T for which the quantities q, λ, and n are defined for all real ξ.) As before, let us temporarily assume that ξ is normally distributed with mean zero and variance σ^2, and set ourselves the task of minimizing

$$\frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} W(y, T)\, e^{-(y^2/2\sigma^2)}\, dy = W^*(T, \sigma^2) \tag{2.2}$$

with respect to T In the next paragraph we digress for a moment to derive a needed elementary inequality.

Let us prove that, if h, h_1, and h_2 are non-negative, and

$$h^2 = p\, h_1^2 + (1 - p)\, h_2^2, \tag{2.3}$$

where $0 < p < 1$, then

$$M(h) \leq p\, M(h_1) + (1 - p)\, M(h_2). \tag{2.4}$$

Hold h and p fixed. The desired result is obviously true when $h_1 = h_2 = h$. Let h_1 and h_2 vary, subject to (2.3). Then

$$\frac{dh_2}{dh_1} = \frac{-ph_1}{(1 - p)h_2}.$$

Also

$$\frac{p\,dM(h_1)}{dh_1} = \frac{-p}{\sqrt{2\pi}}\, e^{-\frac{1}{2}h_1^2}$$

and

$$(1 - p)\, \frac{dM(h_2)}{dh_1} = (1 - p)\, \frac{dM(h_2)}{dh_2}\, \frac{dh_2}{dh_1} = \frac{ph_1}{\sqrt{2\pi}h_2}\, e^{-\frac{1}{2}h_2^2}.$$

Thus the derivative of the right member of (2.4) with respect to h_1 is 0 when $h_1 = h$, positive when $h_1 > h$, and negative when $h_1 < h$. From this we get (2.4).

Let T be any estimation procedure and $L_T(x_1, \cdots, x_n)$ its associated function. Write

$$l_T(x_1, \cdots, x_n) = L_T(x_1, \cdots, x_n) - \bar{x}\left[1 + \frac{1}{n\sigma^2}\right]^{-1}.$$

If $n = m$ and $x_1, \cdots, x_m$ is the sample obtained, we have that the conditional expected value of $W^*(T, \sigma^2)$ is

$$(2.5) \qquad M\left(l_T(x_1, \cdots, x_m)\sqrt{m + \frac{1}{\sigma^2}}\right) + cm + kE(U_m^* + l_T(x_1, \cdots, x_m))^2,$$

where U_m^* is a normally distributed chance variable with mean zero and variance $\left(m + \frac{1}{\sigma^2}\right)^{-1}$. The last term in (2.5) is therefore

$$k\left[\left(m + \frac{1}{\sigma^2}\right)^{-1} + l_T^2(x_1, \cdots, x_m)\right]$$

This is an even function of l_T, while the first term of (2.5) is a monotonically decreasing function of l_T. Thus (2.5) and hence $W^*(T, \sigma^2)$ will be minimized by taking l_T non-negative. Now take the expected value of (2.5) over the set of samples where $n = m$. Application of the result of the preceding paragraph to the finite sums which approximate the integral gives the result that $W^*(T, \sigma^2)$ is minimized when $l_T(x_1, \cdots, x_m)$ is a function only of m. Hence we may restrict ourselves to consideration of procedures T for which (2.5) takes the value

$$(2.6) \qquad M\left(\sqrt{m + \frac{1}{\sigma^2}}\, l_T(m)\right) + cm + k\left[\left(m + \frac{1}{\sigma^2}\right)^{-1} + \{l_T(m)\}^2\right].$$

For any such procedure T, since k and c are fixed positive numbers (and σ^2 is held fixed for the present), the expression (2.6) takes its minimum for some value of m. Thus, in our quest for a procedure T which will minimize $W^*(T, \sigma^2)$ we may restrict ourselves to procedures of *fixed* sample size. This fixed sample size and the (constant) value of l_T are functions of k, c, and σ^2. For fixed m,

$$M\left(\sqrt{m + \frac{1}{\sigma^2}}\, l^0\right) + k(l^0)^2$$

has an absolute minimum at l_m, say, since it is a continuous function of $l^0 (l^0 \geq 0)$ which approaches ∞ with l^0. The case $m = 0$ must be considered. (In this event $\bar{x} = 0$.) Now consider the sequence

$$\left\{M\left(\sqrt{m + \frac{1}{\sigma^2}}\, l_m\right) + cm + k\left[\left(m + \frac{1}{\sigma^2}\right)^{-1} + l_m^2\right]\right\}$$

for $m = 0, 1, 2, \cdots$ ad inf. This sequence condenses only at ∞. Hence there exists a value $N(k, c, \sigma^2)$ of m for which the elements of this sequence have a minimum value We may choose $N(k, c, \sigma^2)$ so that $\lim_{\sigma^2=\infty} N(k, c, \sigma^2)$ exists. (We verify easily that this is always possible.) Designate this limit by $N(k, c, \infty)$,

and the associated l by $l(k, c, \infty)$. The l associated with $N(k, c, \sigma^2)$ will be designated by $l(k, c, \sigma^2)$. Thus a best procedure for minimizing $W^*(T, \sigma^2)$ is to take the fixed number $N(k, c, \sigma^2)$ observations, and to use, as upper bound for ξ, the quantity

$$\bar{x}\left[1 + \frac{1}{\sigma^2 N(k, c, \sigma^2)}\right]^{-1} + l(k, c, \sigma^2).$$

We see readily that

$$l(k, c, \infty) = \lim_{\sigma^2 \to \infty} l(k, c, \sigma^2)$$

and that

$$M(\sqrt{N(k, c, \infty)}\, l(k, c, \infty)) = \lim_{\sigma^2 \to \infty} M\left(\sqrt{N(k, c, \sigma^2) + \frac{1}{\sigma^2}}\; l(k, c, \sigma^2)\right).$$

Let $T(\sigma^2)$ be the procedure described above which is a best procedure T in the sense of minimizing $W^*(T, \sigma^2)$ when σ^2 is the variance of ξ.

We now compute $W(\xi, T(\sigma^2))$ and obtain

$$\text{(2.7)}\qquad \begin{aligned} W(\xi, T(\sigma^2)) = cN + k\left[\frac{N\sigma^4}{(1 + N\sigma^2)^2} + \left(l - \frac{\xi}{1 + N\sigma^2}\right)^2\right] \\ + M\left(\frac{1 + N\sigma^2}{\sqrt{N}\sigma^2}\left[l - \frac{\xi}{1 + N\sigma^2}\right]\right), \end{aligned}$$

where for brevity we have written N and l for $N(k, c, \sigma^2)$ and $l(k, c, \sigma^2)$. Let

$$l - \frac{\xi}{1 + N\sigma^2} = x, \qquad \frac{1 + N\sigma^2}{\sqrt{N}\sigma^2} = \sqrt{N} + \epsilon.$$

Then

$$\text{(2.8)}\qquad W = cN + k\left[\frac{1}{(\sqrt{N} + \epsilon)^2} + x^2\right] + M([\sqrt{N} + \epsilon]\, x),$$

$$\text{(2.9)}\qquad \frac{\partial W}{\partial x} = 2kx - \frac{(\sqrt{N} + \epsilon)}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\{(\sqrt{N} + \epsilon)^2 x^2\}\right].$$

The second term above is always of the same sign and the exponential decreases as $|x|$ increases. Thus $\partial W/\partial x = 0$ has the unique positive root x^*. Put x^* for x in W (in 2.8) and call the result W^*. W is a continuous function of x and approaches ∞ as $|x| \to \infty$. Since the root x^* is unique it follows that W^* is the minimum value of W with respect to x. Now $N(k, c, \sigma^2)$ is constant for σ^2 sufficiently large. Hence, for such σ^2, we have

$$\text{(2.10)}\qquad \begin{aligned} \frac{\partial W^*}{\partial \epsilon} &= \frac{-2k}{(\sqrt{N} + \epsilon)^3} + 2kx^* \frac{dx^*}{d\epsilon} - \frac{dx^*}{d\epsilon}\frac{(\sqrt{N} + \epsilon)}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\{(\sqrt{N} + \epsilon)^2 x^{*2}\}\right] \\ &\quad - \frac{x^*}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\{(\sqrt{N} + \epsilon)^2 x^{*2}\}\right] \\ &= \frac{-2k}{(\sqrt{N} + \epsilon)^3} - \frac{x^*}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\{(\sqrt{N} + \epsilon)^2 x^{*2}\}\right] \end{aligned}$$

since x^* is the root of $\partial W/\partial x = 0$. Also ϵ is positive and, for σ^2 sufficiently large, approaches zero monotonically as σ^2 approaches ∞. For $\epsilon > 0$ we have that $\partial W^*/\partial\epsilon < 0$, since $x^* > 0$. We conclude: For σ^2 sufficiently large,

$$\min_{\xi} W(\xi, T(\sigma^2))$$

increases monotonically with σ^2 and approaches

$$cN + k\left[\frac{1}{N} + \{x_N(k)\}^2\right] + M(\sqrt{N}\,x_N(k)),$$

where N is short for $N(k, c, \infty)$ and $x_N(k)$ is the unique positive root of the equation in x

$$2kx = \frac{\sqrt{N}}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}Nx^2\right]$$

Going back to the definition of $l(k, c, \infty)$ we see that the latter satisfies the equation in l:

$$\frac{d}{dl}\{M(\sqrt{N}\, l) + kl^2\} = 0.$$

Hence

$$x_N(k) = l(k, c, \infty)$$

Thus the classical estimation procedure C_0 where one takes the fixed number $N(k, c, \infty)$ of observations and uses as upper bound for the mean $\bar{x} + l(k, c, \infty)$ is a minimax procedure T, i.e.,

$$W(\xi, C_0) = \inf_T \sup_\xi W(\xi, T)$$

For fixed N, $x_N(k)$ decreases monotonically from $+\infty$ to 0 as k increases from 0 to $+\infty$. Hence, for given positive integral N_0 and $l^* > 0$, there is a unique positive value k_0 such that $x_{N_0}(k_0) = l^*$. Consider the expression

$$B(m) = M(\sqrt{m}\, x_m(k_0)) + cm + k_0\left[\frac{1}{m} + \{x_m(k_0)\}^2\right], \tag{2.11}$$

where m is a positive, continuous variable. We have

$$\begin{aligned}\frac{dB(m)}{dm} = c - \frac{k_0}{m^2} + \frac{dx_m(k_0)}{dm}\frac{\partial}{\partial x_m(k_0)}\left\{M(\sqrt{m}\, x_m(k_0)) + k_0[x_m(k_0)]^2\right\}\\ + \frac{\partial M(\sqrt{m}\, x_m(k_0))}{\partial m}.\end{aligned} \tag{2.12}$$

The third term of the right member is identically zero because

$$2k_0 x_m(k_0) = \frac{\sqrt{m}}{\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}m[x_m(k_0)]^2\right\}. \tag{2.13}$$

Further we have

$$\frac{d^2B(m)}{dm^2} = \frac{2k_0}{m^3} - \frac{d}{dm}\,\frac{m^{-\frac{1}{2}}x_m(k_0)}{2\sqrt{2\pi}}\,e^{-\frac{1}{2}m x_m^2(k_0)} \tag{2.14}$$

$$= \frac{2k_0}{m^3} - \frac{k_0\,d\{m^{-1}(x_m(k_0))^2\}}{dm}.$$

For typographic simplicity we shall use y for $x_m(k_0)$ in the computations of the next few lines. From (2.13) we obtain

$$\log 2k_0 + \log y = -\log \sqrt{2\pi} + \tfrac{1}{2}\log m - \tfrac{1}{2}\, m\, y^2,$$

$$\frac{1}{y}\frac{dy}{dm} = \frac{1}{2m} - \frac{y^2}{2} - my\frac{dy}{dm},$$

$$\frac{dy}{dm} = \frac{y(1 - my^2)}{2m(1 + my^2)}.$$

Hence

$$\frac{d^2B(m)}{dm^2} = 2k_0 m^{-3} + k_0 m^{-2}\, y^2 - 2k_0 m^{-1}\, y\frac{dy}{dm}$$

$$= 2k_0 m^{-3} + k_0 m^{-2}\, y^2 - \frac{k_0 y^2(1 - my^2)}{m^2(1 + my^2)} \tag{2.15}$$

$$= 2k_0 m^{-3} + \frac{2y^4 k_0}{m(1 + my^2)} > 0.$$

Since $c > 0$, we have

$$\lim_{m=0} B(m) = \lim_{m=\infty} B(m) = +\infty.$$

Hence there exists a value of m for which $B(m)$ takes its minimum value. If in $d\,B\,(m)/dm$ we put $m = N_0$ and set the resulting expression equal to zero, we obtain an equation in c whose unique solution c_0, if it is positive, assures us that, when $c = c_0$ and $k = k_0$, $B\,(m)$ takes its minimum at $m = N_0$. A simple computation gives

$$c_0 = \frac{k_0}{N_0^2} + \frac{l^* \exp\{-\tfrac{1}{2}N_0 l^{*2}\}}{2\sqrt{2\pi N_0}} > 0. \tag{2.16}$$

Actually we are interested in considering $B\,(m)$ only for positive integral values of m. We see readily that the minimum of $B\,(m)$ occurs then at $m = N_0$ when c is such that

$$c_1\,(N_0\,,\,k_0) \le c \le c_2\,(N_0\,,\,k_0), \tag{2.17}$$

with c_1 and c_2 roots of the following equations in c:

$$B\,(N_0) = B\,(N_0 + 1),$$

$$B\,(N_0) = B\,(N_0 - 1).$$

(If $N_0 = 1$, then $c_2 = \infty$.)

Let $C_0(N_0, l^*)$ be the classical (non-sequential) procedure where one takes N_0 observations and uses $\bar{x} + l^*$ as upper bound for the mean. Choose $k = k_0$ and c such that (2.17) is satisfied. Then

$$W(\xi, C_0(N_0, l^*)) = cN_0 + k_0\left(\frac{1}{N_0} + l^{*2}\right) + M(\sqrt{N_0}\, l^*)$$

identically in ξ. $C_0(N_0, l^*)$ is a procedure T such that

$$W(\xi, C_0) = \inf_T \sup_\xi W(\xi, T). \tag{2.18}$$

Whenever c and k are given, the N and l of the minimax solution may be obtained as follows: First we obtain an integer N such that

$$c_1(N, k) \leq c \leq c_2(N, k).$$

Knowing N and k we can then solve for l.

The results of this section may be summarized as follows: For every positive c and k there exists a classical estimation procedure $C_0(N, l)$ with positive integral N and $l > 0$ such that (2.18) holds. Conversely, for every such pair (N, l) there exists a positive pair (c, k) so that (2.18) holds. A method of finding one member of the pair of couples (c, k) and (N, l) when the other is given, has been indicated above.

Let T_1 be any procedure for giving an upper bound for ξ. We shall say that T_1 is optimum if for any other procedure T_2 such that

$$\sup_\xi q(\xi, T_2) \leq \sup_\xi q(\xi, T_1),$$

$$\sup_\xi \lambda(\xi, T_2) \leq \sup_\xi \lambda(\xi, T_1),$$

we have

$$\sup_\xi n(\xi, T_2) \geq \sup_\xi n(\xi, T_1).$$

It is easy to prove that the classical procedure C_0 with any positive l and positive integral N is optimum by using the results of the last paragraph. For let $1 - \alpha = M(l\sqrt{N})$ and let k and c be the corresponding parameters. We have then

$$\sup_\xi q(\xi, T_2) + k \sup_\xi \lambda(\xi, T_2) + c \sup_\xi n(\xi, T_2) \geq \sup_\xi \{q(\xi, T_2)$$

$$+ k\, \lambda(\xi, T_2) + cn(\xi, T_2)\} \geq (1 - \alpha) + k\left(\frac{1}{N} + l^2\right) + cN.$$

Since $\sup q(\xi, T_2) \leq (1 - \alpha)$ and $\sup \lambda(\xi, T_2) \leq 1/N + l^2$, we must have

$$\sup_\xi n(\xi, T_2) \geq N,$$

which is the desired result.

In a general unprecise way we may say that an estimation procedure is the better the smaller the three quantities

$$\beta_1(T) = \sup_\xi q(\xi, T), \qquad \beta_2(T) = \sup_\xi \lambda(\xi, T), \qquad \beta_3(T) = \sup_\xi n(\xi, T).$$

We can now assert the following: No sequential procedure T can be superior to the classical fixed sample procedure C in the sense that

$$\beta_i(T) \leq \beta_i(C) \qquad \text{for } i = 1, 2, 3$$

and the inequality sign holds for at least one i.

In concluding this section we may remark that the case $\alpha \leq \frac{1}{2}$, i.e., $l \leq 0$, may be handled in the same manner as above except that we use $M(-l\sqrt{m})$ in place of $M(l\sqrt{m})$.

3. Miscellaneous results; point estimation. Without going into the necessarily involved details, we content ourselves with pointing out that the problem of estimating sequentially the mean of a normal distribution by a finite interval of length not specified in advance, can be solved in similar fashion. As before let ξ be the unknown mean of a normal distribution with unit variance, where ξ may be any real value. We want to estimate by an interval

$$(L_1(x_1, \cdots, x_n), \qquad L_2(x_1, \cdots, x_n)).$$

Let c, k_1, and k_2 be positive constants and consider the problem of minimizing the supremum with respect to ξ of

$$1 - P\{L_1 < \xi < L_2 \mid G^1\} + cn(\xi, G^1) + k_1 E[(L_1 - \xi)^2 \mid G^1] + k_2 E[(L_2 - \xi)^2 \mid G^1],$$

where G^1 is the generic designation of the estimation procedure. As before, employ an a priori normal distribution of ξ with mean zero and variance σ^2, and let $\sigma^2 \to \infty$. A fixed sample size procedure will be a minimax solution. It will possess optimum properties similar to those described in the preceding sections. The problem of minimizing the supremum with respect to ξ of

$$1 - P\{L_1 < \xi < L_2 \mid G^1\} + cn(\xi, G^1) + kE\{(L_2 - L_1)^2 \mid \xi, G^1\}$$

can be treated similarly.

Suppose the sample size is fixed in advance. The problem of finding an estimate which will minimize

$$\sup_\xi [1 - P\{L_1 < \xi < L_2 \mid G^1\} + k_1 E\{(L_1 - \xi)^2 \mid G^1\} + k_2 E\{(L_2 - \xi)^2 \mid G^1\}]$$

or

$$\sup_\xi [1 - P\{L_1 < \xi < L_2 \mid G^1\} + kE\{(L_2 - L_1)^2 \mid \xi, G^1\}]$$

can be treated by the method of the preceding sections.

The problem of estimating (sequentially or with fixed sample size) the means of a multivariate normal distribution with known covariance matrix can be treated in similar fashion.

Suppose it is desired to estimate sequentially the mean ξ $(-\infty < \xi < \infty)$ of a normal distribution with unit variance by means of a chance point

$\hat{\xi}(x_1, \cdots, x_n)$. Let $R(\xi, \xi^1)$ be the Wald risk function (cf. [2]), a non-negative function which measures the loss incurred in using the particular value ξ^1 as an estimate when ξ is the actual value. The functions $\hat{\xi}(x_1, \cdots, x_n)$ and $R(\xi, \xi^1)$ must have suitable measurability properties for which we refer the reader to [2]. Let us seek a procedure ξ^* such that

$$\sup_{\xi}[E\{R(\xi, \xi^*)\} + cn(\xi, \xi^*)] = \inf_{\hat{\xi}} \sup_{\xi}[E\{R(\xi, \hat{\xi})\} + c\, n(\xi, \hat{\xi})].$$

Here $n(\xi, \hat{\xi})$ is the average number of observations under $\hat{\xi}$ when ξ is the "true" mean. The procedure ξ^* will be called a minimax solution. We shall assume that $R(a, b)$ is a monotonically non decreasing function of $| a - b |$, and that there exists a positive number g such that

$$\int_0^{\infty} R(0, x) \exp\left\{-\frac{x^2}{2g}\right\} dx < \infty .$$

As examples of functions with these properties we may cite

$$R(a, b) = | a - b |,$$

$$R(a, b) = (a - b)^2.$$

As before, assume temporarily that ξ is normally distributed with mean zero and variance σ^2. We verify without difficulty that a solution $\hat{\xi} = \xi_0$ which minimizes

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} [E\{R(\xi, \hat{\xi})\} + cn(\xi, \hat{\xi})] \exp\left\{-\tfrac{1}{2} \frac{\xi^2}{\sigma^2}\right\} d\xi$$

is the following: n is identically a suitable constant, say N, and ξ_0 is $\bar{x}(1 + 1/N\sigma^2)^{-1} = \bar{x}h$ say, so that $h < 1$. For this solution we have

$$E\{R(\xi, \xi_0)\} + cn(\xi, \xi_0) = cN + \frac{\sqrt{N}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R(\xi, \bar{x}h) \exp\left\{-\frac{N}{2}(\bar{x} - \xi)^2\right\} d\bar{x}.$$

Write $u = \bar{x} - \xi$. Then

$$R(\xi, \bar{x}h) = R(\xi, h[\xi + u]) = R(0, hu - [1 - h]\xi),$$

$$\int_{-\infty}^{\infty} R(\xi, \bar{x}h) \exp\left\{-\frac{N}{2}(\bar{x} - \xi)^2\right\} d\bar{x}$$

$$= \int_{-\infty}^{\infty} R(0, hu - [1 - h]\xi) \exp\left\{-\frac{Nu^2}{2}\right\} du$$

$$= \int_{-\infty}^{\infty} R(0, v) \exp\left\{-\frac{N}{2h^2}(v + [1 - h]\xi)^2\right\} \frac{1}{h}\, dv.$$

Because of the assumptions on the function R the last expression is a minimum when $\xi = 0$. We may always choose N such that, for large enough σ^2, the integer N is a constant, say N_0. Also $h \to 1$ as $\sigma^2 \to \infty$. Thus we conclude that the follow-

ing is a minimax solution: $n = N_0$ and $\hat{\xi} = \xi^* = \bar{x}$. If any estimation procedure $\hat{\xi}$ is such that $\sup_\xi n(\xi, \hat{\xi}) \leq N_0$ then

$$\sup_\xi E\{R(\xi, \hat{\xi})\} \geq E\{R(\xi, \xi^*)\}.$$

If $\hat{\xi}$ is such that

$$\sup_\xi E\{R(\xi, \hat{\xi})\} < E\{R(\xi, \xi^*)\},$$

then

$$\sup_\xi n(\xi, \hat{\xi}) \geq N_0 .$$

If the restrictions imposed above on R are satisfied and if the sample must always be of given size N, the above argument still holds when $1 \leq N \leq g$, and shows that the estimate $\bar{x}$ minimizes

$$\sup_\xi E\{R(\xi, \hat{\xi})\}$$

with respect to $\hat{\xi}$.

REFERENCES

[1] C. Stein and A. Wald, "Sequential confidence intervals for the mean of a normal distribution with known variance," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 427–433.

[2] A. Wald, "Statistical decision functions," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 165–205.

ASYMPTOTIC PROPERTIES OF THE WALD-WOLFOWITZ TEST OF RANDOMNESS

By Gottfried Emanuel Noether

New York University

1. Summary. The paper investigates certain asymptotic properties of the test of randomness based on the statistic $R_h = \sum_{i=1}^{n} x_i x_{i+h}$ proposed by Wald and Wolfowitz. It is shown that the conditions given in the original paper for asymptotic normality of R_h when the null hypothesis of randomness is true can be weakened considerably. Conditions are given for the consistency of the test when under the alternative hypothesis consecutive observations are drawn independently from changing populations with continuous cumulative distribution functions. In particular a downward (upward) trend and a regular cyclical movement are considered. For the special case of a regular cyclical movement of known length the asymptotic relative efficiency of the test based on ranks with respect to the test based on original observations is found. A simple condition for the asymptotic normality of R_h for ranks under the alternative hypothesis is given. This asymptotic normality is used to compare the asymptotic power of the R_h-test with that of the Mann T-test in the case of a downward trend.

2. Introduction. The hypothesis of randomness, i.e., the assumption that the chance variables $X_1, \cdots, X_n$ have the joint cumulative distribution function (*cdf*) $F(x_1, \cdots, x_n) = F(x_1) \cdots F(x_n)$ where $F(x)$ may be any *cdf*, is basic in many statistical problems. Several tests of randomness designed to detect changes in the underlying population have been suggested, however mostly on intuitive grounds. Very seldom has the actual performance of a test with respect to a given class of alternatives been investigated. It is the intention of this paper to carry out such an investigation for the particular test based on the statistic

$$R_h = \sum_{i=1}^{n} x_i x_{i+h}, \qquad x_{n+j} = x_j,$$

proposed by Wald and Wolfowitz [1]. It is suggested in [1] that this test is suitable if the alternative to randomness is the existence of a trend or a regular cyclical movement. Both these cases will be treated.

Let $a_1, \cdots, a_n$ be observations on the chance variables $X_1, \cdots, X_n$ and assume that the hypothesis of randomness is true. (Henceforth this hypothesis will be denoted by H_0 while the hypothesis that an alternative to randomness is true will be denoted by H_1.) Restricting then $X_1, \cdots, X_n$ to the subpopulation of permutations of $a_1, \cdots, a_n$, any one of the $n!$ possible permutations is equally likely, and the distribution of R_h in this subpopulation can be found. If

the level of significance α is chosen in such a way that $\alpha = m/n!$ where m is a positive integer, the test is performed by selecting m of the $n!$ possible values of R_h and rejecting H_0 when the actually observed value of R_h is one of these m values. The particular choice of these m values should be such as to maximize the power of the test with respect to the class of alternatives under consideration.

Denote the expected value and variance of R_h in the population of equally likely permutations of n observations $a_1, \cdots, a_n$ by E^0R_h and V^0R_h, respectively. Then it is shown in [1] that if h is prime to n

$$E^0 R_h = \frac{1}{n-1}(A_1^2 - A_2) \tag{2.1}$$

and

$$\begin{aligned} V^0 R_h = {} & \frac{1}{n-1}(A_2^2 - A_4) \\ & + \frac{1}{(n-1)(n-2)}(A_1^4 - 4A_1^2A_2 + 4A_1A_3 + A_2^2 - 2A_4) \\ & - \frac{1}{(n-1)^2}(A_1^2 - A_2)^2, \end{aligned} \tag{2.2}$$

where $A_r = a_1^r + \cdots + a_n^r$, $(r = 1, 2, 3, 4)$. Actually (2.1) and (2.2) are valid as soon as $n > 2h$.

Let $R_h^0 = (R_h - E^0R_h)/\sqrt{V^0R_h}$. Then it is also shown in [1] that if h is prime to n, R_h^0 is asymptotically normally distributed with mean 0 and variance 1 provided the a_i, $(i = 1, \cdots, n)$, satisfy condition W:

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(a_i - \bar{a})^r}{\left[\frac{1}{n}\sum_{i=1}^{n}(a_i - \bar{a})^2\right]^{r/2}} = O(1),^1 \qquad (r = 3, 4, \cdots),$$

where $\bar{a} = n^{-1}\sum_{i=1}^{n} a_i$.

It is easily seen that condition W is satisfied when the original observations are replaced by ranks. When the $a_1, \cdots, a_n$ are independent observations on the same chance variable X, condition W is satisfied with probability 1 provided X has positive variance and finite moments of all orders. It is interesting to compare this condition for asymptotic normality of R_h in the population of permutations of observations on the chance variable X with the condition for asymptotic normality of R_h under random sampling. For this case Hoeffding and Robbins [3] have shown that it is sufficient to assume that X has a finite absolute moment of order 3. Thus it is desirable to weaken condition W. This will be done in Section 3.

In further sections the consistency and efficiency of the test based on R_h will

---

[1] The symbol O, as well as the symbols o and $\sim$ to be used later, have their usual meaning. See, for example, Cramér [2], p. 122

be examined assuming that under the alternative hypothesis observations, though still independent, are drawn from changing populations. Throughout the paper the circularly defined statistic R_h is used However, if with probability 1

$$x_{n-h+1}x_1 + \cdots + x_n x_h = o(R_h),$$

it is seen that asymptotically the test based on the non-circular

$$\bar{R}_h = \sum_{i=1}^{n-h} x_i x_{i+h}$$

has the same properties as that based on R_h. We find

$$E^0 \bar{R}_h = \frac{n-h}{n(n-1)} (A_1^2 - A_2),$$

$$V^0 \bar{R}_h = \frac{n-h}{n(n-1)} (A_2^2 - A_4) + \frac{2(n-2h)}{n(n-1)(n-2)} (A_1^2 A_2 - A_2^2 - 2A_1 A_3 + 2A_4)$$
$$+ \frac{(n-h-1)(n-h-2) + 2(h-1)}{n(n-1)(n-2)(n-3)} (A_1^4 - 6A_1^2 A_2 + 8A_1 A_3 + 3A_2^2 - 6A_4)$$
$$- \frac{(n-h)^2}{n^2(n-1)^2} (A_1^2 - A_2)^2.$$

3. Asymptotic normality of R_h under randomization. Let the set of chance variables $X_1, \cdots, X_n$ be defined on the $n!$ equally likely permutations of n numbers $\mathfrak{A}_n = (a_1, \cdots a_n)$. Then we have

THEOREM 1: *The distribution of R_h^0 tends to the normal distribution with mean 0 and variance 1 as $n \to \infty$ provided*

$$(3.1) \qquad \frac{\sum_{i=1}^{n} (a_i - \bar{a})^r}{\left[\sum_{i=1}^{n} (a_i - \bar{a})^2\right]^{r/2}} = o[n^{(2-r)/4}], \qquad (r = 3, 4, \cdots),$$

where $\bar{a} = n^{-1} \sum_{i=1}^{n} a_i$.

REMARK: The set $\mathfrak{A}_n$ need not be a subset of $\mathfrak{A}_{n+1}$.

The proof of this theorem will be omitted, since it is very similar to the proof of another theorem by the author [4].

THEOREM 2: *If the $a_1, a_2, \cdots$ are independent observations on a chance variable X having positive variance and a finite absolute moment of order $4 + \delta$, $\delta > 0$, condition* (3.1) *is satisfied unless possibly an event of probability* 0 *has occurred*

The proof of this theorem will be based on Markoff's method for proving the central limit theorem in the Liapounoff form.[2] Thus we shall show that there exists a sequence of sequences $\mathfrak{B}_n = (b_{n1}, \cdots, b_{nn})$ such that unless possibly an event of probability 0 has occured, (i) there exists an index n' (depending

[2] See, for example, Uspensky [5], pp. 388–95.

on the given sequences such that for $n > n'$, $\mathfrak{A}_n = \mathfrak{B}_n$, and (ii) the sequences $\mathfrak{B}_n$ satisfy condition (3.1) expressed in terms of the b_{ni}, ($i = 1, \cdots, n$).

It is no restriction to assume that $EX = 0$, since the addition of one and the same constant to every a_i does not change (3.1). Let

$$N = N(n) = n^{1/(4+\delta/2)},$$

and define for $i = 1, \cdots, n$

$$\begin{aligned} b_{ni} &= a_i, & c_{ni} &= 0, & &\text{if } a_i \leq N(n), \\ &= 0, & &= a_i, & &\text{if } a_i > N(n), \end{aligned}$$

so that $a_i = b_{ni} + c_{ni}$. Then b_{ni} and c_{ni} can be considered as observations on chance variables Y_n and Z_n, respectively, where

$$\begin{aligned} Y_n &= X, & Z_n &= 0, & &\text{if } X \leq N(n), \\ &= 0, & &= X, & &\text{if } X > N(n). \end{aligned}$$

Further let $p_n = P\{Z_n = X\}$, $\alpha_r(U) = EU^r$, $\beta_r(U) = E|U|^r$ where $U = X$, Y_n, Z_n and r is positive integral, if these moments exist, $\beta_{4+\delta} = E|X|^{4+\delta}$, and finally, let $F(x)$ be the *cdf* of X.

In order to prove (i) consider the infinitely dimensional sample space Ω with the generic point $\omega = \omega(a_1, a_2, \cdots)$ and let $E_n = \{\omega : a_n > N(n)\}$, ($n = 1, 2, \cdots$). Then E_n has probability measure p_n. We shall show that $\sum_{n=1}^{\infty} p_n$ converges. Since

$$\beta_{4+\delta} = \int_{-\infty}^{\infty} |x|^{4+\delta}\, dF(x) \geq N^{4+\delta}\left[\int_{-\infty}^{-N} dF(x) + \int_{+N}^{\infty} dF(x)\right] \geq N^{4+\delta} p_n,$$

we find

$$p_n \leq \beta_{4+\delta} \frac{1}{N^{4+\delta}} = \beta_{4+\delta} \frac{1}{n^{(4+\delta)/(4+\delta/2)}}.$$

Now $(4 + \delta)/(4 + \delta/2) > 1$ and the infinite sum converges. It follows that the set E of points which belong to infinitely many sets E_n has probability measure 0. Thus for every point $\omega \in \Omega$ except those in a set of measure 0 there exists an index n_ω (depending on ω) such that for $n > n_\omega$

$$a_n \leq N(n). \tag{3.2}$$

Further, since n_ω is finite and $N(n) \to \infty$, it follows that for these points there exists a second index $n'_\omega \geq n_\omega$ such that in addition to (3.2) $a_n \leq N(n'_\omega)$, ($n = 1, \cdots, n_\omega$). Thus except on a set of measure 0 the sequences $\mathfrak{B}_n$ are identical with the sequences $\mathfrak{A}_n$ for $n > n'_\omega$. This proves (i).

In proving (ii) let $B_{nr} = \sum_{i=1}^{n} b_{ni}^r$, ($n, r = 1, 2, \cdots$). We first note that under the assumptions of the theorem $n^{-1}A_r \to \alpha_r(X)$ for $r = 1, 2, 3, 4$ except on a set of measure 0. Thus except on a set of measure 0

$$\bar{a} = n^{-1}A_1 = o(1), \qquad A_2 = \Omega(n),[3] \qquad A_3 = O(n), \qquad A_4 = O(n),$$

[3] A function $f(n)$ is said to be of order $\Omega(n^k)$, k real, if $f(n) = O(n^k)$ and $\liminf_n |f(n)/n^k| > 0$

and therefore by the argument used in proving (i) again except on a set of measure 0

$$\bar{b}_n = n^{-1}B_{n1} = o(1), \qquad B_{n2} = \Omega(n), \qquad B_{n3} = O(n), \qquad B_{n4} = O(n).$$

It follows that in order to prove (ii) it is sufficient to show that

$$B_{nr} = o[n^{(r+2)/4}], \qquad (r = 5, 6, \cdots), \tag{3.3}$$

except on a set of measure 0.

Now for $r \geq 5$

$$\alpha_r(Y_n) \leq \beta_r(Y_n) \leq N^{r-4}\beta_4(Y_n) \leq N^{r-4}\beta_4(X),$$

and therefore

$$\alpha_r(Y_n) = O(N^{r-4}) = O[n^{(r-4)/(4+\delta/2)}].$$

It follows that

$$EB_{nr} - n\alpha_r(Y_n) = O[n^{(r+\delta/2)/(4+\delta/2)}]$$

and

$$\operatorname{var} B_{nr} = n \operatorname{var} Y_n^r = n[\alpha_{2r}(Y_n) - \alpha_r^2(Y_n)] = O[n^{(2r+\delta/2)/(4+\delta/2)}],$$

so that

$$\sigma(B_{nr}) = O[n^{(r+\delta/4)/(4+\delta/2)}].$$

Assume now that for some $r \geq 5$ (3.3) is not satisfied on a set F_r having measure $\epsilon_r > \epsilon > 0$. We shall show that this assumption leads to a contradiction, and that therefore (3.3) is true.

Choose e such that

$$1/2 < e < (16 + r\delta)/(32 + 4\delta). \tag{3.4}$$

Since $r \geq 5$, (3.4) can always be satisfied. Then the infinite sum $\sum_{n=1}^{\infty} (1/n^{2e})$ converges, and a positive constant d can be found in such a way that

$$p = \frac{1}{d^2} \sum_{n=1}^{\infty} \frac{1}{n^{2e}} < \epsilon.$$

If we then write the Tchebysheff inequality

$$P\{| B_{nr} - EB_{nr} | > dn^e\sigma(B_{nr})\} \leq 1/d^2n^{2e},$$

it is seen that except on a set having at most measure p

$$B_{nr} = O\{\max[n^{(r+\delta/2)/(4+\delta/2)}, \qquad n^e n^{(r+\delta/4)/(4+\delta/2)}]\}.$$

Now for $r \geq 5$

$$(r + \delta/2)/(4 + \delta/2) < r/4$$

and by (3.4)

$$\begin{aligned} e + (r + \delta/4)/(4 + \delta/2) &= e + r/4 + (\delta/4 - r\delta/8)/(4 + \delta/2) \\ &< r/4 + (16 + 2\delta)/(32 + 4\delta) = (r + 2)/4, \end{aligned}$$

so that the measure of the set I_n [illegible] equal to ϵ. This contradicts our assumption, thus proving Theorem 2.

4. Consistency. To prove consistency of tests based on permutations of observations $a_1, \cdots, a_n$ the following procedure can be applied. Let the test statistic be $S_n = S(x_1, \cdots, x_n)$ and denote by $E_n^0 = E^0(a_1, \cdots, a_n)$ and $V_n^0 = V^0(a_1, \cdots, a_n)$ the expected value and variance of S_n under the assumption that the set of random variables $X_1, \cdots, X_n$ is restricted to the subpopulation consisting of the $n!$ equally likely permutations of the observations. Assume that for the alternatives under consideration large values of S_n are critical. Then we reject the null hypothesis whenever $(S_n - E_n^0)/\sqrt{V_n^0} > k$ where k is some positive constant depending on the limiting distribution of S_n under the assumption of equally likely permutations and the level of significance. Thus in order to prove consistency we have to show that

$$\lim_{n\to\infty} P\left\{\frac{S_n - E_n^0}{\sqrt{V_n^0}} > k \mid H_1\right\} = 1. \tag{4.1}$$

(4.1) will be satisfied if for some $\epsilon > 0$

$$\lim_{n\to\infty} P\left\{\frac{S_n - E_n^0}{\sqrt{nV_n^0}} > \epsilon \mid H_1\right\} = 1.$$

Thus we shall have proved consistency, if we can show that when H_1 is true, $E_n^0/\sqrt{nV_n^0}$ converges in probability to 0 and there exists some $\epsilon > 0$ such that $\lim_{n\to\infty} P\{S_n/\sqrt{nV_n^0} > \epsilon \mid H_1\} = 1$.

Applying this method to our problem and noting that a corresponding procedure could have been used in the case when small values of S_n are critical, we obtain

THEOREM 3: *The test based on R_h is consistent with respect to alternatives for which*

$$\frac{E^0 R_h}{\sqrt{nV^0 R_h}} \xrightarrow{pr} 0 \tag{4.2}$$

and there exists some $\epsilon > 0$ such that

$$\lim_{n\to\infty} P\left\{\left|\frac{R_h}{\sqrt{nV^0 R_h}}\right| > \epsilon\right\} = 1, \tag{4.3}$$

where E^0R_h and V^0R_h are given by (2.1) *and* (2.2), *respectively.*

In what follows it will always be assumed that under the alternative hypothesis observations are independent from chance variables X_n with continuous *cdf's* $F_n(x)$, $(n = 1, 2, \cdots)$. We shall often have the opportunity to make use of the fact that the test is not changed if one and the same constant is subtracted from every observation. This will be helpful in reducing our problem to one for which (4.2) is true.

Let a_i be the rank of the observation x_i on the chance variable X_i, $(i =$

$1, \cdots, n$). Then it is no restriction to assume that these ranks take the special form

$$-(n-1)/2, \quad -(n-3)/2, \cdots, (n-1)/2,$$

so that $A_1 = 0$, $A_2 = \frac{1}{12}(n^2-1)n = \Omega(n^3)$ and

$$V^0 R_h \sim \frac{1}{n} A_2^2 \sim \frac{1}{144} n^5 = \Omega(n^5) \tag{4.4}$$

and therefore (4.2) is always satisfied.

Before we can find conditions under which (4.3) is satisfied, we have to investigate the expected value and variance of R_h when H_1 is true. For this purpose write $a_i = \sum_{j=1}^n y_{ij}$, $(i = 1, \cdots, n)$,

$$y_{ij} = -1/2 \quad \text{if } x_i > x_j, \qquad y_{ii} = 0. \tag{4.5}$$
$$\phantom{y_{ij}} = 1/2 \quad \text{if } x_i < x_j,$$

Then if $P\{X_i < X_j\} = p_{ij}$, $(i, j = 1, \cdots, n)$, we find

$$Ey_{ij} = \tfrac{1}{2}p_{ij} - \tfrac{1}{2}(1 - p_{ij}) = p_{ij} - \tfrac{1}{2} = \epsilon_{ij}, \qquad \text{(say)}.$$

Further,

$$R_h = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n y_{ij} y_{i+h,k}, \qquad y_{n+l,k} = y_{lk}. \tag{4.6}$$

Therefore

$$E(R_h \mid H_1) = \sum_i \sum_j \sum_k \epsilon_{ij}\epsilon_{i+h,k} + O(n^2) \tag{4.7}$$

and

$$\begin{aligned} \operatorname{var} R_h &= E \sum_{ijk} \sum_{\alpha\beta\gamma} y_{ij} y_{i+h,k} y_{\alpha\beta} y_{\alpha+h,\gamma} - E \sum_{ijk} y_{ij} y_{i+h,k} E \sum_{\alpha\beta\gamma} y_{\alpha\beta} y_{\alpha+h,\gamma} \\ &= \sum_{ijk} \sum_{\alpha\beta\gamma} (Ey_{ij} y_{i+h,k} y_{\alpha\beta} y_{\alpha+h,\gamma} - Ey_{ij} y_{i+h,k} E y_{\alpha\beta} y_{\alpha+h,\gamma}). \end{aligned} \tag{4.8}$$

In (4.8) the expression in parentheses is 0 unless one of the Greek indices (including $\alpha + h$) equals one of the Roman indices. Therefore $\operatorname{var}(R_h \mid H_1) = O(n^5)$.

It then follows from (4.4) that

$$R_h / \sqrt{n V^0 R_h} \sim \frac{12}{n^3} R_h \underset{pr}{\rightarrow} 12 \lim_{n \to \infty} \frac{1}{n^3} E(R_h | H_1),$$

and we can state the following corollary to Theorem 3:

Corollary: *When using ranks, the test based on R_h is consistent, if under the alternative hypothesis*

$$\frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \epsilon_{ij}\epsilon_{i+h,k} = \Omega(1), \tag{4.9}$$

where $\epsilon_{ij} = P\{X_i < X_j\} - \frac{1}{2}$.

Since $\epsilon_{ij} = -\epsilon_{ji}$, we can write

$$\sum_i \sum_j \sum_k \epsilon_{ij}\epsilon_{i+h,k} = \sum_k \sum_i \sum_{j>i} \epsilon_{ij}(\epsilon_{i+h,k} - \epsilon_{j+h,k}) = L, \quad \text{(say)},$$

and the test is consistent if

$$\lim_{n\to\infty} \frac{1}{n^3} L \neq 0. \tag{4.10}$$

4.1. *Downward (upward) trend.* Assume that for $i < j$ and all k

$$\epsilon_{ij} < 0 \tag{4.11}$$

and

$$\epsilon_{ik} \leq \epsilon_{jk}. \tag{4.12}$$

These requirements are equivalent to $P\{X_i < X_j\} < 1/2$ and $P\{X_i < X_k\} \leq P\{X_j < X_k\}$ and are satisfied if the alternative to randomness is a downward trend in the sense that $F_i(x) \leq F_j(x)$, $(-\infty < x < \infty, i < j)$, with at least one interval of strict inequality.

(4.11) and (4.12) are not sufficient for (4.10) to be true. Thus assume in addition that there exist a positive integer n' and a number $\epsilon > 0$ such that l.u.b.$_{j-i\geq n'}\, \epsilon_{ij} = -\epsilon$. Then

$$\lim_{n\to\infty} \frac{1}{n^3} L \geq \lim_{n\to\infty} \frac{1}{n^3} \sum_{k=1}^{n} \sum_{\substack{i \leq k-h-n' \\ j \geq k-h+n'}} \epsilon_{ij}(\epsilon_{i+h,k} - \epsilon_{j+h,k})$$

$$\geq 2\epsilon^2 \lim_{n\to\infty} \frac{1}{n^3} \sum_{k=1}^{n} (k-h-n')(n-k+h-n'+1) = 2\epsilon^2(\tfrac{1}{2} - \tfrac{1}{3}) > 0,$$

and the test is consistent.

The case of an upward trend can be treated in exactly the same way. The test is consistent with respect to alternatives for which for $i < j$ and all k, $\epsilon_{ij} > 0$, $\epsilon_{ik} \geq \epsilon_{jk}$, and g.l.b.$_{j-i\geq n'}\, \epsilon_{ij} = \epsilon$, where this time $\epsilon > 0$.

Another test of randomness, the so-called T-test, has been proposed by Mann [6] with exactly this alternative of a downward (upward) trend in mind. This T-test is also consistent provided certain general conditions are satisfied. Thus the question arises which of the two tests should be chosen if a downward (upward) trend is feared. This question will be considered in Section 7.

4.2. *Cyclical movement.* Let the class of alternatives be specified by

$$\epsilon_{lg+\alpha, mg+\beta} = \epsilon_{\alpha\beta}, \qquad (\alpha, \beta = 1, \cdots, g > 1;\ l, m = 0, 1, \cdots), \tag{4.13}$$

in other words, assume that the statistic R_h is used to test for randomness while under the alternative hypothesis there exists a regular cyclical movement with a period of length g. It is sufficient to consider the case $h \leq g$.

If (4.13) is true,

$$\sum_{i,j,k=1}^{n} \epsilon_{ij}\epsilon_{i+h,k} = n^2 \sum_{i=1}^{n} \epsilon_{i\cdot}\, \epsilon_{i+h,\cdot} + O(n^2) = n^3 \eta + O(n^2), \tag{4.14}$$

where

$$(4.15)\qquad \epsilon_{\alpha.} = \frac{1}{g}\sum_{\beta=1}^{g} \epsilon_{\alpha\beta}$$

and

$$(4.16)\qquad \eta = \frac{1}{g}\sum_{\alpha=1}^{g} \epsilon_{\alpha.}\epsilon_{\alpha+h,.}\,.$$

Thus in view of (4.9) the test is consistent if $\eta \neq 0$.

If $h = g$, η reduces to a sum of squares and is therefore > 0 if some $\epsilon_{\alpha.} \neq 0$. However it is possible that some or even all $\epsilon_{\alpha\beta} \neq 0$, $(\alpha \neq \beta)$, and still $\epsilon_{\alpha.} = 0$. If this happens, the test is inconsistent, otherwise it is consistent. If under H_1 the populations from which consecutive observations are drawn differ only in location, the above mentioned exceptional case cannot happen, and the test is always consistent with respect to this class of alternatives.

If $h \neq g$, it is not difficult to construct an example where $\sum_{\alpha=1}^{g} \epsilon_{\alpha.}\epsilon_{\alpha+h,.} \neq 0$ while $\sum_{\alpha=1}^{g} \epsilon_{r_\alpha.}\epsilon_{r_{\alpha+h}.} = 0$, where the r_α are a permutation of the numbers $1, \cdots, g$. Thus in this case it is not sufficient that some $\epsilon_{\alpha.} \neq 0$ for the test to be consistent. Consistency may also depend on the order of the elements of a period.

We may conclude that if g is known, we should always choose $h = g$. If g is not known, we may as well take $h = 1$.

4.8. Change in location. Turning now to the case when the test is performed on the basis of the original observations, it will often be appropriate to assume that under the alternative hypothesis the distribution remains the same except for a location parameter. We shall consider only the case of a cyclical movement.

Thus let

$$F_n(x) = F(x - m_n) \qquad (n = 1, 2, \cdots),$$

where $F(x)$ is the *cdf* of a chance variable U having mean 0, and m_n is a location parameter. It will also be assumed that U has the positive variance σ^2 and a finite fourth moment.

In the cyclical case with period g

$$(4.17)\qquad m_{lg+\alpha} = m_\alpha \qquad (\alpha = 1, \cdots, g > 1;\ l = 0, 1, \cdots).$$

We shall find conditions under which our test is consistent with respect to alternatives of this kind. Obviously we can assume that $\sum_{\alpha=1}^{g} m_\alpha = g\bar{m} = 0$, since otherwise we could have subtracted $\bar{m}$ from every observation. Writing then $a_n = u_n + m_n$, $(n = 1, 2, \cdots)$, where u_n can be considered as an observation on the previously defined chance variable U, we find

$$A_1 = \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} u_i + O(1),$$

$$A_2 = \sum_{i=1}^{n} u_i^2 + 2\sum_{i=1}^{n} u_i m_i + \sum_{i=1}^{n} m_i^2$$

$$= \sum_{i=1}^{n} u_i^2 + 2\sum_{\alpha=1}^{g} m_\alpha \sum_{l=0}^{n_\alpha} u_{lg+\alpha} + \left[\frac{n}{g}\right]\sum_{\alpha=1}^{g} m_\alpha^2 + O(1),$$

where n_α is the largest integer such that $n_\alpha g + \alpha \leq n$ and $[n/g]$ the largest integer $\leq n/g$. A_3 and A_4 are given by similar expressions. Since we assumed that $EU = 0$, $EU^2 = \sigma^2 > 0$, and $EU^4 < \infty$, we have with probability 1

$$\sum_{i=1}^n u_i = o(n), \qquad \sum_{i=1}^n u_i^2 = \Omega(n), \qquad \sum_{i=1}^n u_i^3 = O(n), \qquad \sum_{i=1}^n u_i^4 = O(n),$$

so that with the same probability

$$A_1 = o(n), \qquad A_2 = \Omega(n), \qquad A_3 = O(n), \qquad A_4 = O(n).$$

It follows that with probability 1

$$E^0 R_h = o(n), \qquad V^0 R_h \sim \frac{1}{n} A_2^2 = \Omega(n),$$

and condition (4.2) of Theorem 3 is satisfied.

Since further

$$\begin{aligned} \operatorname{var} R_h &= \sum_{i=1}^n \operatorname{var}(x_i x_{i+h}) + 2 \sum_{i=1}^n \operatorname{cov}(x_i x_{i+h}, x_{i+h} x_{i+2h}) \\ &= \sum_{i=1}^n \{(\sigma^2 + m_i^2)(\sigma^2 + m_{i+h}^2) - m_i^2 m_{i+h}^2\} \\ &\qquad + 2 \sum_{i=1}^n \{m_i m_{i+2h}(\sigma^2 + m_{i+h}^2) - m_i m_{i+h}^2 m_{i+2h}\} \\ &= \sum_{i=1}^n \{\sigma^4 + \sigma^2(m_i^2 + m_{i+h}^2 + 2m_i m_{i+2h})\} = O(n) \end{aligned} \tag{4.18}$$

and therefore except on a set of probability measure 0

$$\frac{R_h}{\sqrt{nV^0 R_h}} \sim \frac{R_h}{A_2} = \frac{\frac{1}{n} R_h}{\frac{1}{n} A_2} \xrightarrow[pr]{} \frac{\lim_{n\to\infty} \frac{1}{n} E(R_h \mid H_1)}{\sigma^2 + \frac{1}{g} \sum_{\alpha=1}^{g} m_\alpha^2},$$

condition (4.3) is satisfied provided $\lim_{n\to\infty} n^{-1} E(R_h \mid H_1) \neq 0$. Now $E(R_h \mid H_1) = [n/g] \sum_{\alpha=1}^{g} m_\alpha m_{\alpha+h} + O(1)$, so that the test is consistent with respect to the class of alternatives (4.17) for which

$$\sum_{\alpha=1}^{g} (m_\alpha - \bar{m})(m_{\alpha+h} - \bar{m}) \neq 0,$$

where $\bar{m} = g^{-1} \sum_{\alpha=1}^{g} m_\alpha$. Thus by the same argument as in the case of ranks, the test is consistent whenever $h = g$, while it may or may not be consistent if $h < g$.

5. Limiting distribution of R_h under H_1 in case of ranks. For the remaining two sections, it is of importance to know conditions under which R_h based on ranks is asymptotically normal under the alternative hypothesis. Using the methods of moments, it can be shown that in this case the distribution of

$(R_h - ER_h)/\sigma(R_h)$ tends to the normal distribution with mean 0 and variance 1 provided var $R_h = \Omega(n^5)$.

Generalizing the method used in Section 4 in evaluating the variance of R_h, it is not difficult to see that $E(R_h - ER_h)^{2s+1} = O(n^{5s+2})$, $(s = 0, 1, \cdots)$. It follows that if var $R_h = \Omega(n^5)$, the odd moments are asymptotically zero. By means of a more careful analysis, it is also possible to show that $E(R_h - ER_h)^{2s} \sim (2s-1)(2s-3) \cdots 3(\text{var } R_h)^s$. This proves our statement.

6. Ranks versus original observations. We have seen in Section 4 that if the alternative hypothesis is characterized by a regular cyclical movement the test based on R_h is consistent both for original observations and for ranks, provided $h = g$, where g is the length of a cycle. The question arises which test is more efficient, the one based on original observations or the one based on ranks.

In trying to answer this question, we shall make use of a procedure due to Pitman[4], which allows us to compare two consistent tests of the hypothesis that some population parameter θ has the value θ^0 against the alternatives $\theta > \theta^0$ using critical regions of size α, $S_{in} \geq S_{in}(\alpha)$, $(i = 1, 2)$, where S_{in} is a statistic having finite variance and $S_{in}(\alpha)$ is an appropriate constant. The relative efficiency of the second test with respect to the first test is defined as the ratio n_1/n_2 where n_2 is the sample size of the second test required to achieve the same power for a given alternative as is achieved by the first test using a sample of size n_1 with respect to the same alternative.

Let $E(S_{in} \mid \theta) = \psi_{in}(\theta)$, $\text{var}(S_{in} \mid \theta) = \sigma_{in}^2(\theta)$, and $\psi'_{in}(\theta^0)/\sigma_{in}(\theta^0) = H_i(n)$ Assuming that the alternative is of the form $\theta_n = \theta^0 + k/\sqrt{n}$ where k is a positive constant, Pitman has shown that the asymptotic relative efficiency of the second test with respect to the first test is given by $\lim_{n\to\infty} [H_2^2(n)/H_1^2(n)]$, provided there exists a number $\epsilon > 0$ such that for $\theta^0 \leq \theta \leq \theta^0 + \epsilon$

$$\psi'_{in}(\theta) \text{ exists;} \tag{6.1}$$

as $\theta_n \to \theta^0$ with $n \to \infty$

$$\frac{\psi'_{in}(\theta_n)}{\psi'_{in}(\theta^0)} \to 1 \tag{6.2}$$

and

$$\frac{\sigma_{in}(\theta_n)}{\sigma_{in}(\theta^0)} \to 1; \tag{6.3}$$

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} H_i(n) = c_i \text{, where } c_i \text{ is some positive constant;} \tag{6.4}$$

(6.5) the distribution of $[S_{in} - \psi_{in}(\theta)]/\sigma_{in}(\theta)$ tends to the normal distribution with mean 0 and variance 1 uniformly in θ.

[4] I should like to thank Professor Pitman for his kind permission to quote from his lectures on non-parametric statistical inference which he delivered at Columbia University during the spring semester 1948.

Condition(6.5) can be replaced by the weaker condition

(6.5′) the distribution of $[S_{in} - \psi_{in}(\theta_n)]/\sigma_{in}(\theta_n)$ tends to the normal distribution with mean 0 and variance 1 as $n \to \infty$.

In our case, in order to insure consistency, it will be assumed that $h = g$. Consider the parameter

$$\theta = \frac{1}{h} \sum_{\alpha=1}^{h} (m_\alpha - \bar{m})^2, \tag{6.6}$$

where as before m_α is the expected value of the $(lh + \alpha)$th observation, $(l = 0, 1, \cdots)$. We want to find the asymptotic relative efficiency of the test performed on ranks with respect to the test performed on original observations as $\theta \to 0$ with $n \to \infty$.

Again it is no restriction to assume that

$$\bar{m} = \frac{1}{h} \sum_{\alpha=1}^{h} m_\alpha = 0. \tag{6.7}$$

Assume further that the chance variable U defined in 4.3 has a finite absolute moment of order $4 + \delta$, $\delta > 0$. Then $R_h^0 \sim \sqrt{n} R_h / A_2$ with probability 1 and, if the null hypothesis is true, it follows from Theorem 2 that with the same probability the statistic

$$Q_h = \frac{\sqrt{n} \sum_{i=1}^{n} x_i x_{i+h}}{\sum_{i=1}^{n} x_i^2}$$

has in the population of permutations of the observed sample values an asymptotically normal distribution with mean 0 and variance 1. This, however, is also the limiting distribution of Q_h under random sampling when the null hypothesis is true, as follows from the results of Hoeffding and Robbins [3]. Thus it will be sufficient to find the asymptotic relative efficiency of the R_h-test for ranks with respect to the Q_h-test. In doing this, it will also be assumed that U has a continuous density function $f(x) = F'(x)$, and, in order to simplify notation, that there are nh observations instead of n.

In finding $H_Q(nh)$, let $x_{\alpha,j} = x_{\alpha j} = x_{(j-1)h+\alpha}$ and $u_{\alpha,j} = u_{\alpha j} = u_{(j-1)h+\alpha}$, $(\alpha = 1, \cdots, h, j = 1, \cdots, n)$ Then

$$\begin{aligned} \frac{1}{nh} A_2 &= \frac{1}{nh} \sum_{\alpha=1}^{h} \sum_{j=1}^{n} x_{\alpha j}^2 = \frac{1}{nh} \sum_{\alpha=1}^{h} \sum_{j=1}^{n} (u_{\alpha j} + m_\alpha)^2 \\ &= \frac{1}{nh} \sum_{\alpha=1}^{h} \left\{ \sum_{j=1}^{n} u_{\alpha j}^2 + 2m_\alpha \sum_{j=1}^{n} u_{\alpha j} + nm_\alpha^2 \right\} \xrightarrow[pr1]{} \sigma^2 + \theta \end{aligned}$$

Further,

$$R_h = \sum_{\alpha=1}^{h} \left\{ \sum_{j=1}^{n} u_{\alpha j} u_{\alpha,j+1} + 2m_\alpha \sum_{j=1}^{n} u_{\alpha j} + nm_\alpha^2 \right\}$$

so that

$$EQ_h = E\frac{\frac{1}{\sqrt{nh}}R_h}{\frac{1}{nh}A_2} \sim \frac{\sqrt{nh}\,\theta}{\sigma^2+\theta} = \psi_{Qn}(\theta).$$

Therefore

$$\psi'_{Qn}(\theta) = \sqrt{nh}\,\frac{\sigma^2}{(\sigma^2+\theta)^2}.$$

Also by (4.18)

$$\operatorname{var} Q_h \sim \frac{nh\sigma^4 + 4n\sigma^2\sum_{\alpha=1}^{h} m_\alpha^2}{nh(\sigma^2+\theta)^2} = \frac{\sigma^4+4\sigma^2\theta}{(\sigma^2+\theta)^2}$$

which converges to 1 as $\theta \to 0$. It follows that

$$H_Q(nh) = \psi'_{Qn}(0) = \frac{\sqrt{nh}}{\sigma^2}. \tag{6.8}$$

Conditions (6.1)-(6.5) are easily seen to be satisfied.

Considering now the R_h-test for ranks, we know that $(nh)^{-5/2}R_h$ has finite variance. From (4.7) and (4.11)-(4.16) it is found that

$$E[(nh)^{-5/2}R_h \mid \theta] \sim \sqrt{nh}\,\eta = \sqrt{nh}\,\frac{1}{h^3}\sum_{\alpha=1}^{h}\left(\sum_{\beta=1}^{h}\epsilon_{\alpha\beta}\right)^2 = \psi_{Rn}(\theta) \tag{6.9}$$

and after some computations

$$\psi'_{Rn}(0) = \sqrt{nh}\left[\int_{-\infty}^{\infty} f^2(x)\,dx\right]^2. \tag{6.10}$$

From (4.4) and (6.10)

$$H_R(nh) = 12\sqrt{nh}\left[\int_{-\infty}^{\infty} f^2(x)\,dx\right]^2.$$

Conditions (6.1)-(6.4) and (6.5′) can be shown to be satisfied.

Thus the asymptotic relative efficiency of the test based on ranks with respect to the test based on original observations is

$$H_{RQ} = \frac{144nh\left[\int_{-\infty}^{\infty} f^2(x)\,dx\right]^4}{nh/\sigma^4} = 144\left[\sigma\int_{-\infty}^{\infty} f^2(x)\,dx\right]^4. \tag{6.11}$$

As is not difficult to see, this expression is independent of location and scale.

Let the chance variable U have density function

$$f(x) = \begin{cases} 0, & x < -1, \quad x > 1, \\ \dfrac{1+x}{1+a}, & -1 \le x \le a, \\ \dfrac{1-x}{1-a}, & a \le x \le 1, \end{cases} \qquad -1 \le a \le 1,$$

i.e., let the graph of $f(x)$ be given by the two straight lines connecting the points $(-1, 0)$ and $(1, 0)$ with the point $(a, 1)$. Then $EU = a/3$, $\text{var } U = \frac{1}{18}(3 + a^2)$, $\int_{-\infty}^{\infty} f^2(x)\, dx = 2/3$, and (6.11) becomes $[8(3 + a^2)/27]^2$. Thus H_{RQ} increases with $|a|$. For $a = 0$, it is equal to $64/81$; for $|a| = 1$, it is equal to $(32/27)^2$. It is equal to 1, for $a = \sqrt{3/8}$.

This example shows that the asymptotic relative efficiency of the rank test with respect to the test based on original observations may be <1, $=1$, or >1, depending on the density function $f(x)$. Unless $f(x)$ is explicitly given, no statement can be made as to which of the two tests is to be preferred.

We are now in a position to give at least a partial answer to a question raised in [1]. In concluding their paper, Wald and Wolfowitz note that the problem dealt with in this section can be posed not only when transforming to ranks, but also for any transformation carried out by means of a continuous and strictly monotonic function $h(x)$.

Let $t = h(x)$ be such a transformation, satisfying in addition the condition that Pitman's procedure remains applicable for the transformed distribution. Corresponding to σ^2 and Q we shall use σ_t^2 and Q_t. Let $h(m_\alpha) = \mu_\alpha$, $h^{-1} \sum_{\alpha=1}^{h} (\mu_\alpha - \bar{\mu})^2 = \vartheta$. Then if $EQ_t \sim \psi_{Q_t n}(\theta)$, by (6.8), (6.9), and (6.10)

$$\begin{aligned}\frac{d\psi_{Q_t n}(0)}{d\theta}\bigg|_{\theta=0} &= \frac{d\psi_{Q_t n}}{d\vartheta}\frac{d\vartheta}{d\eta}\frac{d\eta}{d\theta}\bigg|_{\theta=0} \\ &= \frac{\sqrt{nh}}{\sigma_t^2}\frac{1}{\left\{\int_{-\infty}^{\infty} f^2[g(t)]g'^2(t)\, dt\right\}^2}\left[\int_{-\infty}^{\infty} f^2(x)\, dx\right]^2 = H_{Q_t}(nh),\end{aligned} \tag{6.12}$$

where $g(t)$ is the inverse of $h(x)$. Therefore by (6.8) and (6.12)

$$H_{Q_t Q} = \frac{\left\{\sigma \int_{-\infty}^{\infty} f^2(x)\, dx\right\}^4}{\left\{\sigma_t \int_{-\infty}^{\infty} f^2[g(t)]g'^2(t)\, dt\right\}^4},$$

and the asymptotic relative efficiency does not merely depend on $h(x)$, the operator defining the transformation, but also very essentially on the underlying distribution $f(x)$.

7. Comparison of the R_h- and T-tests. The T-test by Mann [6] designed to test for randomness against a downward trend is based on the statistic

$$T = \sum_{i=1}^{n} \sum_{j>i} (y_{ij} + \tfrac{1}{2}) = \sum_{i} \sum_{j>i} y_{ij} + \tfrac{1}{4}n(n-1),$$

where y_{ij} is defined by (4.5). Making the same assumptions as in 4.1, Mann shows that under the null hypothesis T has a limiting normal distribution with

mean $\frac{1}{4}n(n-1)$ and variance $\frac{1}{72}(2n^3+3n^2-5n)$, while under the alternative hypothesis

$$ET = \tfrac{1}{4}n(n-1)(2\zeta_n+1), \tag{7.1}$$

where ζ_n is defined by $\frac{1}{2}n(n-1)\zeta_n = \sum_i \sum_{j>i} \epsilon_{ij} < 0$.

Let

$$S_n = \frac{6}{n^{3/2}}\left[T - \tfrac{1}{4}n(n-1)\right].$$

When H_0 is true, S_n is asymptotically normal with mean 0 and variance 1. If we then put $\phi(\lambda) = \frac{1}{\sqrt{2\pi}}\int_\lambda^\infty e^{-\frac{1}{2}x^2}\,dx$, a critical region for testing H_0 is given by $S_n \le -\lambda$, where λ is determined in such a way that $\phi(\lambda) = \alpha$, the level of significance.

When H_1 is true, we find from (7.1)

$$E(S_n \mid \zeta_n) \sim 3\sqrt{n}\,\zeta_n\,.$$

By paralleling the proof of asymptotic normality of R_h under H_1 given in Section 5, it can be shown that $(S_n - ES_n)/\sigma(S_n)$ is asymptotically normal with mean 0 and variance 1 provided $\sigma(S_n) = \Omega(1)$. This is essentially the result obtained already by Hoeffding [7]. Thus the asymptotic power of the test based on S_n is given by

$$P\{S_n \le -\lambda\} \sim \phi\left(\frac{\lambda + 3\sqrt{n}\,\zeta_n}{\sigma(S_n)}\right) \tag{7.2}$$

converging to 1, provided $\lim_{n\to\infty}\sqrt{n}\,\zeta_n = -\infty$. This is the condition for consistency given by Mann.

We may ask for the asymptotic power of the S_n-test as $\zeta_n \to 0$ with $n \to \infty$. More exactly, instead of considering a certain alternative $\epsilon_{ij} = k_{ij}$, where the k_{ij} are given constants, consider the alternative (changing with n)

$$\epsilon_{ij} = \frac{k_{ij}}{\sqrt{n}}. \tag{7.3}$$

If then as $n \to \infty$

$$\frac{2}{n(n-1)}\sum_i \sum_{j>i} k_{ij} \to k$$

and

$$\sigma(S_n) \to 1,$$

it follows from (7.2) that the asymptotic power of the S_n-test, and therefore of the T-test, for alternatives (7.3) is equal to

$$\phi(\lambda + 3k).$$

Now consider the same situation when the statistic R_h is used instead of T. We know that when H is true

$$R'_n = \frac{12}{n^{3/2}} R_h ,$$

where R_h is given by (4.6), is asymptotically normal with mean 0 and variance 1. Thus in this case the critical region is given by $R'_n > \lambda$. If we set $\xi_n = \frac{1}{n^2}\sum$ [illegible], we find

$$E(R'_n \mid \xi_n) \sim 12\sqrt{n}\,\xi_n ,$$

and asymptotically the power of the R'_n-test is

$$(7.4) \qquad P\{R'_n > \lambda\} \sim \varphi\left(\frac{\lambda - 12\sqrt{n}\,\xi_n}{\sigma(R'_n)}\right),$$

provided $\sigma(R'_n) = \Omega(1)$. Thus the test is consistent if $\lim_{n\to\infty} \sqrt{n}\,\xi_n = \infty$. However, for the alternative (7.3), (7.4) tends to $\varphi(\lambda) = \alpha$, provided that as $n \to \infty$

$$\sigma(R'_n) \to 1.$$

Thus the R_h test is ineffective with respect to the alternative (7.3) in contrast to the T-test. This means that for this alternative the asymptotic relative efficiency of the R_h-test with respect to the T-test is 0.

Acknowledgment. The author wishes to acknowledge the valuable help of Professor J. Wolfowitz who suggested the topic and under whose direction the work was completed.

REFERENCES

[1] A. Wald and J. Wolfowitz, "An exact test for randomness in the non-parametric case, based on serial correlation," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 378-388.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, Princeton, 1946.

[3] W. Hoeffding and H. Robbins, "The central limit theorem for dependent random variables," *Duke Math. J.*, Vol. 15 (1948), pp. 773-780.

[4] G. E. Noether, "On a theorem by Wald and Wolfowitz," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 455-458.

[5] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937.

[6] H. B. Mann, "Nonparametric tests against trend," *Econometrica*, Vol. 13 (1945), pp. 245-259.

[7] W. Hoeffding, "A class of statistics with asymptotically normal distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 293-325.

THE DISTRIBUTION OF THE NUMBER OF EXCEEDANCES[1]

By E. J. Gumbel and H. von Schelling

New York and Naval Medical Research Laboratory, New London, Connecticut

0. The problem. We study the probability that the mth observation in a sample of size n taken from an unknown distribution of a continuous variate will be exceeded x times in N future trials, and calculate the averages, the moments, and the cumulative probability function of the number of exceedances. This problem leads to the hypergeometric series. Our starting point is a special case of a distribution studied by Wilks [3] who considered several order statistics whereas we consider only one. His tolerance limits are special cases of our cumulative probability function. Thus the present paper is, at the same time, a specialization and a generalization of the work done by Wilks

1. Distribution. From a continuous variate ξ an alternative is constructed by choosing the mth among n observations $\xi_m(m = 1, 2, \cdots, n)$. The rank m is counted from the top, which means that $m = 1$ $(m = n)$ stands for the largest (smallest) observation. The observation ξ_m is thus the mth largest value. We ask: In how many cases x will the past mth observation be equalled or exceeded in N future trials taken from the same population? For the sake of simplicity, x is called the number of exceedances.

If the initial probability $F(\xi_m) = F_m$ for a value less than ξ_m is known, the alternative probability for exceeding ξ_m is $1 - F_m$, and Bernoulli's theorem gives the probability

$$w_1(F_m, N, x) = \binom{N}{x}(1 - F_m)^x F_m^{N-x} \tag{1.1}$$

that x among N future trials will exceed ξ_m. However, as a rule the probability F_m is unknown. The only data known are the n past observations. To eliminate the probability F_m, we introduce the distribution $v(F_m)$ of the frequency F_m of the *mth* largest among n values

$$v(n, m, F_m)\, dF_m = \binom{n}{m} mF_m^{n-m}(1 - F_m)^{m-1}\, dF_m, \tag{1.2}$$

consider F_m as a variate, and integrate (1.1) over all values of this variate Thus F_m is replaced by a function of n and m.

The convolution of (1.1) and (1.2) leads to *the distribution* $w(n, m, N, x)$ *of*

[1] Opinions or conclusions contained in this paper are those of the authors They are not to be construed as necessarily reflecting the views or endorsement of the Navy Department.

the number of exceedances x over the mth of n past observations in N future trials

$$w(n, m, N, x) = \frac{\binom{n}{m} \cdot m \cdot \binom{N}{x}}{(N+n)\binom{N+n-1}{m+x-1}}. \tag{1.3}$$

This probability depends upon the parameters n, m, and N, but not upon the unknown probability F_m. Therefore it is distribution free. If we are interested in the dependence of $w(n, m, N, x)$ on x only, we simply write $w(x)$. The conditions for the positive integers m and x, and for the probability $w(x)$ are

$$1 \le m \le n; \quad 0 \le x \le N; \quad \sum_{x=0}^{N} w(x) = 1. \tag{1.3'}$$

The distribution (1.3) possesses the following *symmetry*

$$w(n, m, N, x) = w(n, n - m + 1, N, N - x) \tag{1.4}$$

which reads: *The probability that the past mth value from above will be exceeded x times in N new trials is equal to the probability that the past mth value from below will be exceeded N − x times.*

The nN probabilities $w(n, m, N, x)$ are linked by several recurrence formulas which follow easily from the usual combinatorial rules. For fixed m, the probability for $x + 1$ is obtained from the probability for x by

$$\begin{aligned} w(n, m, N, x + 1) &= w(n, m, N, x)\,\frac{(N - x)(m + x)}{(N + n - m - x)(x + 1)} \\ &= w(n, n - m + 1, N, N - x). \end{aligned} \tag{1.5}$$

In the same way, the probabilities $w(n, m, N, x + 1)$, $w(n, m + 1, N, x)$ and $w(n, x, N, m)$ are easily obtained from the probabilities $w(n, m, N, x)$. The distribution (1.3) has many aspects since, besides the number of exceedances x, also the rank m and the number of future trials N may be considered as variates.

For $m = 1$ and $m = n$, the distribution of the number of exceedances over the largest value diminishes with x, and the distribution of the number of exceedances over the smallest value increases with x. For $x = 0$, and $m = 1$, we obtain from (1.3)

$$w(n, 1, N, 0) = \frac{n}{N + n} = w(n, n, N, N). \tag{1.6}$$

For $x = 0$, $m = n$, the probability that the smallest observation will never be exceeded, equal to the probability that the largest value will always be exceeded, is very small, even for moderate sample sizes.

If n is odd, then $m = (n + 1)/2$ corresponds to the median of the initial variable ξ, and the symmetry relation (1.4) becomes

$$w(n, (n + 1)/2, N, x) = w(n, (n + 1)/2, N, N - x). \tag{1.7}$$

It is equally probable that the median of the n past observations is surpassed x or $N - x$ times in N future trials.

2. The two asymptotic distributions. If both n and N are large, m may increase with n such that the quotient m/n remains constant, and the mth values remain near the median. Or, m remains constant such that $m << n$, and the mth values are extremes.

In the first case, let $n = N = 2k - 1$, where k is large. Then $m = k$ is the rank of the median of the initial distribution As shown in (1.7), the distribution of the number of exceedances over the initial median is symmetrical. To obtain the asymptotic distribution we reduce x by writing

$$x = k + z\sqrt{k} \tag{2.1}$$

where z remains in a finite interval The same reduction may be applied to mth values in the neighborhood of the initial median. The distribution of the number of exceedances over the initial median is, from (1.3) and (2.1),

$$w(2k - 1, k, 2k - 1, x) = \text{const}\, \frac{\binom{2k-1}{k + z\sqrt{k}}}{\binom{4k-3}{2k + z\sqrt{k} - 1}}.$$

Consider only the factors involving the variate z, then the right side becomes, by Stirling's formula,

$$\frac{(2k + z\sqrt{k} - 1)!(2k - z\sqrt{k} - 2)!}{(k + z\sqrt{k})!(k - z\sqrt{k} - 1)!} \sim \frac{(2k + z\sqrt{k})^{2k+z\sqrt{k}}\,(2k - z\sqrt{k}]^{2k-z\sqrt{k}}\, e^{-z\sqrt{k}+z\sqrt{k}}}{(k + z\sqrt{k})^{k+z\sqrt{k}}\,(k - z\sqrt{k})^{k-z\sqrt{k}}\, e^{-z\sqrt{k}+z\sqrt{k}}}.$$

Combination of the factors with the same powers leads to

$$\frac{(4k^2 - kz^2)^{2k}}{(k^2 - kz^2)^k}\left(\frac{(2k + z\sqrt{k})\,(k - z\sqrt{k})}{(2k - z\sqrt{k})\,(k + z\sqrt{k})}\right)^{z\sqrt{k}} \sim \frac{\left(1 - \frac{z^2}{4k}\right)^{2k}}{\left(1 - \frac{z^2}{k}\right)^k}\left(\frac{\left(1 + \frac{z}{2\sqrt{k}}\right)\left(1 - \frac{z}{\sqrt{k}}\right)}{\left(1 - \frac{z}{2\sqrt{k}}\right)\left(1 + \frac{z}{\sqrt{k}}\right)}\right)^{z\sqrt{k}}.$$

Since k and $\sqrt{k}$ are large, and z is small, all factors lead to exponential functions whence

$$\exp\left[-\frac{z^2}{2} + z^2 + \frac{z^2}{2} + \frac{z^2}{2} - z^2 - z^2\right] = \exp\left[-\frac{z^2}{2}\right]$$

and finally,

$$\lim_{k\to\infty} w(2k - 1, k, 2k - 1, x) = \text{const}\, e^{-z^2/2}. \tag{2.2}$$

The number of exceedances over the initial median, $m = k$, in a large sample of size $2k - 1$ in $2k - 1$ future trials is normally distributed with mean, median, mode, and variance equal to k. Therefore the probabilities (2 2) may be called *the distribution of normal exceedances.*

In the second case where N and n are large, and m and x are small, a distribution analogous to the Poisson distribution will be obtained. To indicate that N and n are large, they are written $\underline{N}$ and $\underline{n}$. The probability

$$w(\underline{n}, m, \underline{N}, x) = \frac{(x + m - 1)!\,\underline{n}!\,\underline{N}!\,(\underline{N} + \underline{n} - x - m)!}{(m - 1)!\,x!\,(\underline{n} - m)!\,(\underline{N} - x)!\,(\underline{N} + \underline{n})!}$$

obtained from (1 3) becomes, by use of the Stirling formula,

$$w(\underline{n}, m, \underline{N}, x) = \binom{x + m - 1}{x} \frac{\underline{n}^m \underline{N}^x}{(\underline{N} + \underline{n})^{m+x}} \tag{2.3}$$
$$= w(\underline{n}, \underline{n} - m + 1, \underline{N}, \underline{N} - x).$$

If $\underline{n} = \underline{N}$, the preceding formula becomes

$$w(\underline{n}, m, \underline{n}, x) = \binom{x + m - 1}{x}(\tfrac{1}{2})^{m+x} = w(\underline{n}, \underline{n} - m + 1, \underline{n}, \underline{n} - x) \tag{2.4}$$

This probability that the mth largest (or smallest) value will be exceeded x times (or $\underline{n} - x$ times) in $\underline{n}$ future trials is independent of $\underline{n}$. Since m is small compared to $\underline{n}$, the probabilities (2.4) may be called *the distribution of rare exceedances.*

For $x = 0$, we obtain the probability

$$w(\underline{n}, m, \underline{n}, 0) = (\tfrac{1}{2})^m = w(\underline{n}, \underline{n} - m + 1, \underline{n}, \underline{n})$$

that the largest (or smallest) mth extreme value is never (or always) exceeded. For $m = 1$, and $\underline{n} = \underline{N}$, the probability

$$w(\underline{n}, 1, \underline{n}, x) = (\tfrac{1}{2})^{x+1} = w(\underline{n}, \underline{n}, \underline{n}, \underline{n} - x) \tag{2 5}$$

that the largest (or smallest) value is exceeded x times (or $\underline{n} - x$ times) is a geometric series.

To obtain the moments of the distribution of rare exceedances (2.4) we construct its generating function

$$G_x(t) = (\tfrac{1}{2})^m \sum_{x=0}^{\infty} \binom{x + m - 1}{m - 1}\left(\frac{e^t}{2}\right)^x.$$

From the well known expression for the negative binomial follows

$$G_x(t) = (\tfrac{1}{2})^m \left(1 - \frac{e^t}{2}\right)^{-m}, \tag{2 6}$$

whence, by the usual procedure

$$\bar{x} = m \tag{2 7}$$

The mean number of exceedances over the mth value from above in the dis-

tribution of rare exceedances is m itself. The second derivative of (2.6) for $t = 0$ leads to the variance

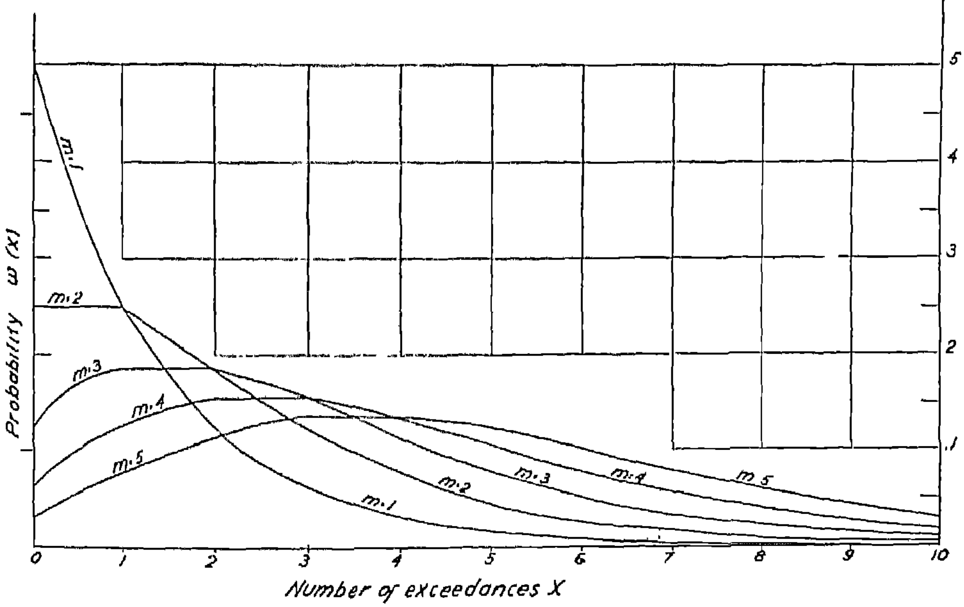
$$\sigma^2 = 2m \tag{2.8}$$

which is the double of the variance in the Poisson distribution. This difference is easily explained: If we apply the Poisson law to the exceedances, we have to know the mean number of exceedances. In our case we only know one observed number of exceedances. Consequently the variance must be larger than in the Poisson case.

GRAPH 1

The distribution of rare exceedances

$w(x) = \binom{x+m-1}{x}\left(\frac{1}{2}\right)^{m+x}$



The variance for the distribution (2 2) of the normal exceedances was $(N + 1)/2$, whereas the variance (2.8) for the distribution of rare exceedances, $2m$, is much smaller since m is small compared to N. This interesting relation will be generalized in paragraph 3.

For m increasing, the distributions (2.4) spread as shown in graph 1. The distributions have two modes

$$\tilde{x}_1 = m - 2, \tilde{x}_2 = m - 1 \tag{2.9}$$

except for $m = 1$, where the probability diminishes with x. The distributions (2 4) are similar to the Poisson distribution for integer m. However, for this distribution the modes are $m - 1$ and m

The similarity between the two distributions may also be seen from their behavior for large m. In this case, the Poisson distribution for the standardized variate $y = (x - m)/\sigma$ converges toward a normal distribution. The same holds for the distribution of rare exceedances. For the proof consider the standardized variate

$$y = (x - m)/\sqrt{2m}. \tag{2.10}$$

Its moment generating function $G_y(t)$ becomes, from (2.6),

$$G_y(t) = \left(2e^{t/\sqrt{2m}} - e^{2t/\sqrt{2m}}\right)^{-m}.$$

The usual development leads to the second member

$$\left(2 + \frac{2t}{\sqrt{2m}} + \frac{2t^2}{4m} - 1 - \frac{2t}{\sqrt{2m}} - \frac{4t^2}{4m} + O(m^{-3/2})\right)^{-m}$$
$$= \left(1 - \frac{t^2}{2m} + O(m^{-3/2})\right)^{-m}.$$

If we neglect the factors $O(m^{-3/2})$, we finally obtain

$$G_y(t) = e^{t^2/2} \tag{2.11}$$

which is the normal generating function. Thus the distribution of rare exceedances converges toward normalcy in the same way as the Poisson distribution.

3. Moments. We return to the general distribution (1.3). For the calculation of the moments, the hypergeometric series $F(\alpha, \beta, \gamma, 1)$ defined by

$$F(\alpha, \beta, \gamma, 1) = 1 + \frac{\alpha}{1}\frac{\beta}{\gamma} + \frac{\alpha(\alpha+1)}{1\cdot 2}\frac{\beta(\beta+1)}{\gamma(\gamma+1)} + \cdots \tag{3.1}$$

is used. The $x + 1$st member of this series is

$$f(x) = \frac{\alpha(\alpha+1)\cdots(\alpha+x-1)}{x!}\frac{\beta(\beta+1)\cdots(\beta+x-1)}{\gamma(\gamma+1)\cdots(\gamma+x-1)}. \tag{3.2}$$

On the other hand, the $x + 1$st member of the distribution $w(x)$ may be written, from (1.3), after changing the signs,

$$w(x) = \frac{\binom{n}{m}}{\binom{N+n}{m}} \frac{m(m+1)\cdots(m+x-1)}{x!} \cdot \frac{(-N)(-N+1)\cdots(-N+x-1)}{(m-n-N)(m-n-N+1)\cdots(m-n-N+x-1)}. \tag{3.3}$$

This is the general member (3.2) of the hypergeometric series, if we write

$$\alpha = m, \beta = -N; \gamma = m - n - N \tag{3.4}$$

Therefore the probability $w(n, m, N, x)$ is the $x + 1$st member in the development of

$$\frac{n!(N + n - m)!}{(N + n)!(n - m)!} F(m, -N, m - n - N, 1).$$

Since the sum of the probabilities $w(x)$ must be unity, we obtain

$$F(m, -N, m - n - N, 1) = \frac{(N + n)!}{n!} \frac{(n - m)!}{(N + n - m)!}. \tag{3.5}$$

This relation will be used for the calculation of the factorial moments $\bar{x}_{[k]}$ of order k which are, from (3.3.),

$$\begin{aligned} \bar{x}_{[k]} &= \frac{n!(N + n - m)!}{(n - m)!(N + n)!} \sum_{x=k}^{N} \\ &\frac{N(N - 1) \cdots (N - x + 1)m(m + 1) \cdots (m + x - 1)}{(x - k)!(N + n - m)(N + n - m - 1) \cdots (N + n - m - x + 1)}. \end{aligned} \tag{3.6}$$

The first member in the sum is

$$\varphi(1) = \frac{N(N - 1) \cdots (N - k + 1)m(m + 1) \cdots (m + k + 1)}{0!(N + n - m)(N + n - m - 1) \cdots (N + n - m - k + 1)}. \tag{3.7}$$

The second member is

$$\varphi(2) = \varphi(1) \frac{(N - k)(m + k)}{1!(N + n - m - k)}.$$

Generally, each successive member is obtained from the preceding one by the same rules as the successive members of the hypergeometric series (3.1) Consequently, from (3.6),

$$\bar{x}_{[k]} = \frac{n!(N + n - m)!}{(n - m)!(N + n)!} \varphi(1) \left(1 + \frac{(N - k)(m + k)}{1!(N + n - m - k)} + \cdots\right). \tag{3 8}$$

The sum in the brackets is the hypergeometric series

$$F(m + k, -(N - k), (m - n - N + k), 1).$$

If we replace, in (3.5), m by $m + k$, N by $N - k$, n by $n + k$,we obtain for the sum in (3.8)

$$\begin{aligned} &F(m + k, -(N - k), m - n - N + k, 1) \\ &\qquad = \frac{(N + n)!(n - m)!}{(n + k)!(N + n - m - k)!}. \end{aligned} \tag{3.9}$$

Introduction of (3 9) and (3.7) into (3.8) leads to the factorial moments

$$\bar{x}_{[k]} = \frac{m(m + 1) \cdots (m + k - 1)N(N - 1) \cdots (N - k + 1)}{(n + 1)(n + 2) \cdots (n + k)} \tag{3.10}$$

and to the recurrent relation

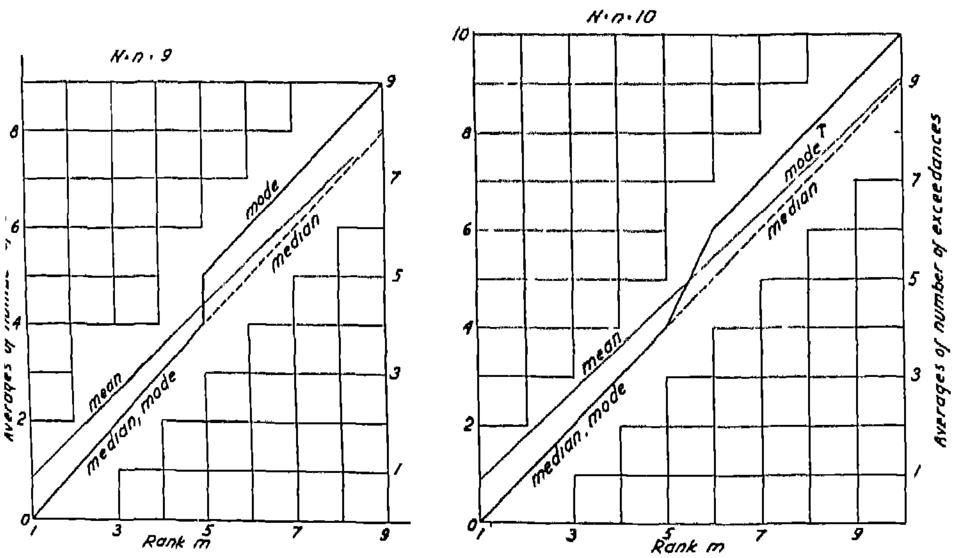
$$\bar{x}_{[k]} = \frac{(m+k-1)(N-k+1)}{n+k}\,\bar{x}_{[k-1]}. \tag{3.10'}$$

If n and N are both of the same order of magnitude, and large compared to k, the expression (3.10) simplifies to

$$\bar{x}_{[k]} = m(m+1)\cdots(m+k-1). \tag{3.10''}$$

GRAPH 2

Averages of numbers of exceedances.



For $k = 1$ we obtain the mean number of exceedances $\bar{x}_m$ over the mth largest value in N future trials

$$\bar{x}_m = N\frac{m}{n+1}. \tag{3.11}$$

This expression is identical with the classical formula $\bar{x} = N(1 - F_m)$ in the Bernoulli distribution (1.1), since the mean of $1 - F_m$ obtained from (1.2) is $m/(n+1)$ In both distributions the means need not be integers. The mean number of exceedances over the smallest value is n times the mean number of exceedances over the largest value. If $N = n + 1$, we have $\bar{x}_m = m$, and the same holds if n and N are large. If n is odd, and $m = (n+1)/2$, the mean number of exceedances over the median of n observations is $N/2$. The means $\bar{x}_m$ are traced against m in Graph 2 for $n = N = 9$, and $n = N = 10$.

The mean number ${}_m\bar{x}$ of exceedances over the mth value from below is related to $\bar{x}_m$ by

$$\bar{x}_m + {}_m\bar{x} = N. \tag{3.12}$$

The variances σ_m^2 and ${}_m\sigma^2$ of the number of exceedances over the mth values from above and below become, from (3.10),

$$\overline{x^2} - \bar{x}^2 = \frac{mN}{n+1}\left(1 + \frac{(m+1)(N-1)}{n+2} - \frac{mN}{n+1}\right).$$

The choice of a common denominator leads, after trivial calculations, to

$$\sigma_m^2 = \frac{mN(n-m+1)(N+n+1)}{(n+1)^2(n+2)} = {}_m\sigma^2 \tag{3.13}$$

The variances increase with N and diminish strongly with increasing n *The variance is maximum* for $m = (n+1)/2$, i.e. *for the median observation* where it becomes

$$\sigma^2_{(n+1)/2} = \frac{N}{4}\frac{(N+n+1)}{(n+2)}. \tag{3.13'}$$

The variances of the number of exceedances over the largest and the smallest value are

$$\sigma_1^2 = \frac{nN(N+n+1)}{(n+1)^2(n+2)} = {}_1\sigma^2. \tag{3.13''}$$

The quotient of the variances of the median and of the extremes is

$$\frac{\sigma^2_{(n+1)/2}}{\sigma_1^2} = \frac{(n+1)^2}{4n} = \frac{\sigma^2_{(n+1)/2}}{{}_1\sigma^2}. \tag{3.14}$$

Consequently the variance of the median is about $n/4$ times larger than the variance of the extremes. In other words, *the extremes are more reliable* than the median, and this quality increases with the sample size. This is a generalization of the relation obtained in paragraph 2. Such a behavior seems singular. However, it also holds for the uniform distribution, and for the distribution (1 2) of the frequencies [1].

In Bernoulli's case, the variance σ_B^2 is, after replacing $1 - F_m$ by $m/(n+1)$,

$$\sigma_B^2 = N\frac{m}{(n+1)}\frac{(n-m+1)}{(n+1)},$$

whence, from (3.13),

$$\sigma_m^2 = \sigma_B^2\frac{N+n+1}{n+2} > \sigma_B^2.$$

The variance in our case is larger than in Bernoulli's case, since we do not assume the knowledge of the probability F_m which is required for the Bernoulli distribu-

tion. For $N = n + 3$, the [illegible] of the Bernoulli distribution. This is a generalization of formula [illegible].

4. The mode and the median. We ask for the most probable number $\tilde{x}$ of exceedances over the previous mth largest among n observations in N future trials. If a mode exists, it must be an integer. Since the distribution $w(x)$ decreases (or increases) with x for $m = 1$ (or $m = n$), we only consider

(4.1) $$2 \leq m \leq n - 1.$$

The mode is obtained from the inequalities

(4.2) $$w(n, m, N, \tilde{x} - 1) \leq w(n, m, N, \tilde{x}) \geq w(n, m, N, \tilde{x} + 1)$$

which lead, from (1.5) to

(4.3) $$(m - 1)\frac{N + 1}{n - 1} - 1 \leq \tilde{x} \leq (m - 1)\frac{N + 1}{n - 1}.$$

The length of the interval is unity, as for the Bernoulli distribution.

There are several cases where *two* modes exist.

a) Let the number of future trials N be such that

(4.4) $$N = k(n - 1) - 1$$

where k is a positive integer. Then the modes are, from (4.3)

(4.5) $$\tilde{x}_{(1)} = k(m - 1) - 1; \tilde{x}_{(2)} = k(m - 1).$$

b) The modes (4.5) also hold if n and N are large compared to unity, and if $N = k'n$, where k' is again an integer.

c) If n is odd, the median of the initial variate has the rank $m = (n + 1)/2$. If, at the same time, N is odd, there are two modes, namely

(4.6) $$\tilde{x}_{(1)} = (N - 1)/2; \tilde{x}_{(2)} = (N + 1)/2$$

In the case $N = n$, the two modes $\tilde{x}_{(1)} = m - 1$, and $\tilde{x}_{(2)} = m$ differ by unity from the modes valid in the two previous cases.

In the case $n = N$, and $m \neq (n + 1)/2$, only one mode exists. To find its location, consider first the case that $n = N$ is even, and $m \leqq n/2$. Then the upper limit in (4.3) is

$$[m - 1] + \frac{2}{n - 1}(m - 1) \leqq [m - 1] + 1 - \frac{1}{n - 1} < [m].$$

Since the interval has unit length, the mode is $\tilde{x} = m - 1$. If $m > (n + 1)/2$, the lower limit is

$$[m - 2] + \frac{2}{n - 1}(m - 1) > [m - 1].$$

The case that $n = N$ is odd is treated in the same way, and leads to the follow-

ing result: The most probable numbers of exceedances over the mth value in $N = n$ future trials are

$$\bar{x} = m - 1 \text{ for } m \leqq n/2; \ \bar{x} = m \text{ for } m > (n/2) + 1, \qquad \text{if } n = N \text{ is even,}$$

$$(4.7) \qquad \bar{x} = m - 1 \text{ for } m \leqq (n + 1)/2; \ \bar{x} = m \text{ for } m \geqq (n + 1)/2, \qquad \text{if } n = N \text{ is odd.}$$

We now consider the median If the probabilities $w(x)$ are summed up from $x = 0$ onward, there may exist an integer $\breve{x}_m$ such that the probability for at most $\breve{x}_m$ exceedances is $\frac{1}{2}$. This is the median number of exceedances. Such a number need not exist. Assume, for example, $N < n$, then the probability $w(n, 1, N, 0)$ alone (see (1.6)) surpasses $\frac{1}{2}$, and the number of exceedances over the largest and the smallest value do not possess a median. If the median $\breve{x}_m$ exists, it follows from the symmetry (1.4) that $N - \breve{x}_m - 1$ is the median of the number of exceedances over the mth value from below. The relation

$$(4.8) \qquad \breve{x}_m + {}_m\breve{x} = N - 1$$

differs from the corresponding relation (3.12) for the mean. In some special cases, the median can be obtained immediately. For $x = 0, m = 1, n = N$, formula (1.6) leads to

$$w(n, 1, n, 0) = \tfrac{1}{2} = w(n, n, n, n).$$

The probability that the largest (or smallest) of n past observations will never (or always) be exceeded in n future trials is equal to $\frac{1}{2}$. If n and N are odd, and $m = (n + 1)/2$, the summation of equation (1.7) yields, with the help of (1.3′),

$$\sum_{0}^{x} w(z) = \sum_{N-x}^{N} w(z) = 1 - \sum_{x+1}^{N} w(z).$$

Now the median number of exceedances $\breve{x}$ is such that the two sums on the right sides are equal to $\frac{1}{2}$. Consequently the median number of exceedances in this case is $m - 1$

We claim that

$$(4.9) \qquad \breve{x}_m = m - 1$$

for all m, provided that $n = N$. For the proof, consider the probability $W(n, m, N, x)$ that the mth largest value is exceeded at most x times in N future trials. This is the sum of the first $x + 1$ members $w(x)$. Let $F_v(\alpha, \beta, \gamma, 1)$ be the sum of the first v members of the hypergeometric series (3 1) Then the substitutions (3.4) and $v = x + 1$ lead to

$$(4.10) \qquad W(n, m, N, x) = \frac{\binom{n}{m}}{\binom{N + n}{m}} F_{x+1}(m, -N, m - n - N, 1).$$

For the sums of the hypergeometric series $F(\alpha, \beta, \gamma, 1)$ the following recurrence formula [2] is used.

$$\frac{(\gamma - \beta - \alpha)(\gamma - \beta - \alpha + 1) \cdots (\gamma - \beta - 1)}{(\gamma - \alpha)(\gamma - \alpha + 1) \cdots (\gamma - 1)} F_\nu(\alpha, \beta, \gamma, 1)$$

$$= 1 - \frac{\beta(\beta + 1) \cdots (\beta + \nu - 1)}{(\gamma - \alpha)(\gamma - \alpha + 1) \cdots (\gamma - \alpha + \nu - 1)} F_\alpha(\nu, \gamma - \beta - \alpha, \gamma - \alpha + \nu, 1). \tag{4.11}$$

The substitutions used in (3.4), and $\nu = x + 1$ lead to

$$\frac{(-n)(-n+1)\cdots(-n+m-1)}{(-n-N)(-n-N+1)\cdots(-n-N+m-1)} F_{x+1}(m, -N, m-n-N, 1)$$

$$= 1 - \frac{(-N)(-N+1)\cdots(-N+x)}{(-n-N)(-n-N+1)\cdots(-n-N+x)} F_m(x+1, -n, -n-N+x+1, 1).$$

This equation may be written from (4.10)

$$W(n, m, N, x) = 1 - \frac{\binom{N}{x+1}}{\binom{N+n}{x+1}} F_m(x+1, -n, -n-N+x+1, 1). \tag{4.12}$$

For $x = m - 1$, and $N = n$, the equation becomes

$$W(n, m, n, m-1) = 1 - \frac{\binom{n}{m}}{\binom{2n}{m}} F_m(m, -n, -2n+m, 1)$$

From (4 10) it follows that the second factor on the right side is equal to the left side

$$W(n, m, n, m-1) = \frac{\binom{n}{m}}{\binom{2n}{m}} F_m(m, -n, -2n+m, 1)$$

Consequently

$$W(n, m, n, m-1) = \tfrac{1}{2} \tag{4.13}$$

If $n = N$, the median number $\check{x}_m$ of exceedances over the mth largest value is $m - 1$, as stated previously The means, modes, and medians obtained from the exact formulae (3 11), (4 7) and (4.9) are traced in graph (2) for $n = N = 9$, and $n = N = 10$

5. Probabilities of at least one exceedance. If we sum up the probabilities $w(x)$ from zero up to a certain x (or from a certain x up to N), we obtain the probabilities $W(x)$ (or $P(x)$) for at most (or at least) x exceedances over the mth past value in N future trials

$$W(x) = \sum_{z=0}^{x} w(z); \qquad P_{(x)} = \sum_{z=x}^{N} w(z) \tag{5.1}$$

where

$$W(x) + P(x-1) = 1; W(x-1) + P(x) = 1.$$

The boundary conditions are

$$W(0) = w(0); W(N) = 1; P(0) = 1; P(N) = w(N).$$

From the symmetry (1.4) it follows that the probability for the mth value from above to be exceeded at most x times is equal to the probability for the mth value from below to be exceeded at least $N - x$ times.

From (5.1) and (1.3) it follows for $m = 1$ (and $m = n$) that the probabilities for the largest (or smallest) among n observations to be exceeded at most once in n future trials converges toward 3/4 (or zero), respectively. If n is large, the probability that the largest value will be exceeded at most x times in n future trials is, by virtue of (2.5),

$$W(\underline{n}, 1, \underline{n}, x) = 1 - (\tfrac{1}{2})^x = P(\underline{n}, \underline{n}, \underline{n}, \underline{n} - x) \tag{5.2}$$

independent of n.

Consider now the probability that the mth largest value will be exceeded *at least once* in N future trials

$$P(n, m, N, 1) = 1 - \frac{n!}{(n-m)!} \frac{(N+n-m)!}{(N+n)!} \tag{5.3}$$

$$= W(n, n - m + 1, N, N - 1)$$

If N and n are large, and m is small, this expression becomes

$$P(\underline{n}, m, \underline{N}, 1) = 1 - \left(\frac{n}{n+N}\right)^m = W(\underline{n}, \underline{n} - m + 1, \underline{N}, \underline{N} - 1)$$

For $m = 1$ and $n = N$, the probability is $\frac{1}{2}$, independent of the size of n.

The least number of exceedances over the smallest value for given probabilities P, called the *tolerance limit*, has been derived by S. S. Wilks [3] A related problem is the following: How many trials N have to be made in order that there is a given probability α for the mth largest value to be exceeded at least once? By virtue of (5.3) we obtain N from

$$\frac{n!(N+n-m)!}{(n-m)!(N+n)!} = 1 - \alpha \tag{5.4}$$

For the largest value $m = 1$, this equation leads to

$$\frac{N}{n} = \frac{1}{1 - \alpha} - 1. \tag{5.5}$$

Of course, N/n increases with α. If n is large, and m remains small, equation (5.4) leads, in first approximation, to

$$\frac{N}{n} = (1 - \alpha)^{-1/m} - 1. \tag{5.6}$$

The quotients N/n as function of α are traced in graph (3). The quotient is plotted vertically against $1/(1 - \alpha)$ plotted horizontally, both in logarithmic scales. The abscissa shows the probability α. The curve for $m = 1$ is exact The corresponding curves for the penultimate and the two preceding values ($m = 2, 3, 4$) are obtained from the approximation (5.6). The graph reads in the following way: The probability that the largest, or second, or third, or fourth value from above are exceeded at least once in $100n$, or $9n$, or $3.6n$, or $2.2n$ future trials is $\alpha = .99$. Inversely, in $4n$ future trials the probability that the largest, or the second, or the third, or fourth extreme value is exceeded at least once is $\alpha = 0.80$, or 0.96, or 0.992, or 0.9984, respectively.

In a similar way we calculate the probabilities that the largest (and penultimate) among n observations is exceeded at least twice in N future trials Let α_2 be this probability. Then we have for the largest value

$$\begin{aligned} 1 - \alpha_2 &= w(n, 1, N, 0) + w(n, 1, N, 1) \\ &= \frac{n}{n + N}\left(1 + \frac{N}{n + N - 1}\right). \end{aligned}$$

For n sufficiently large, the expression simplifies to

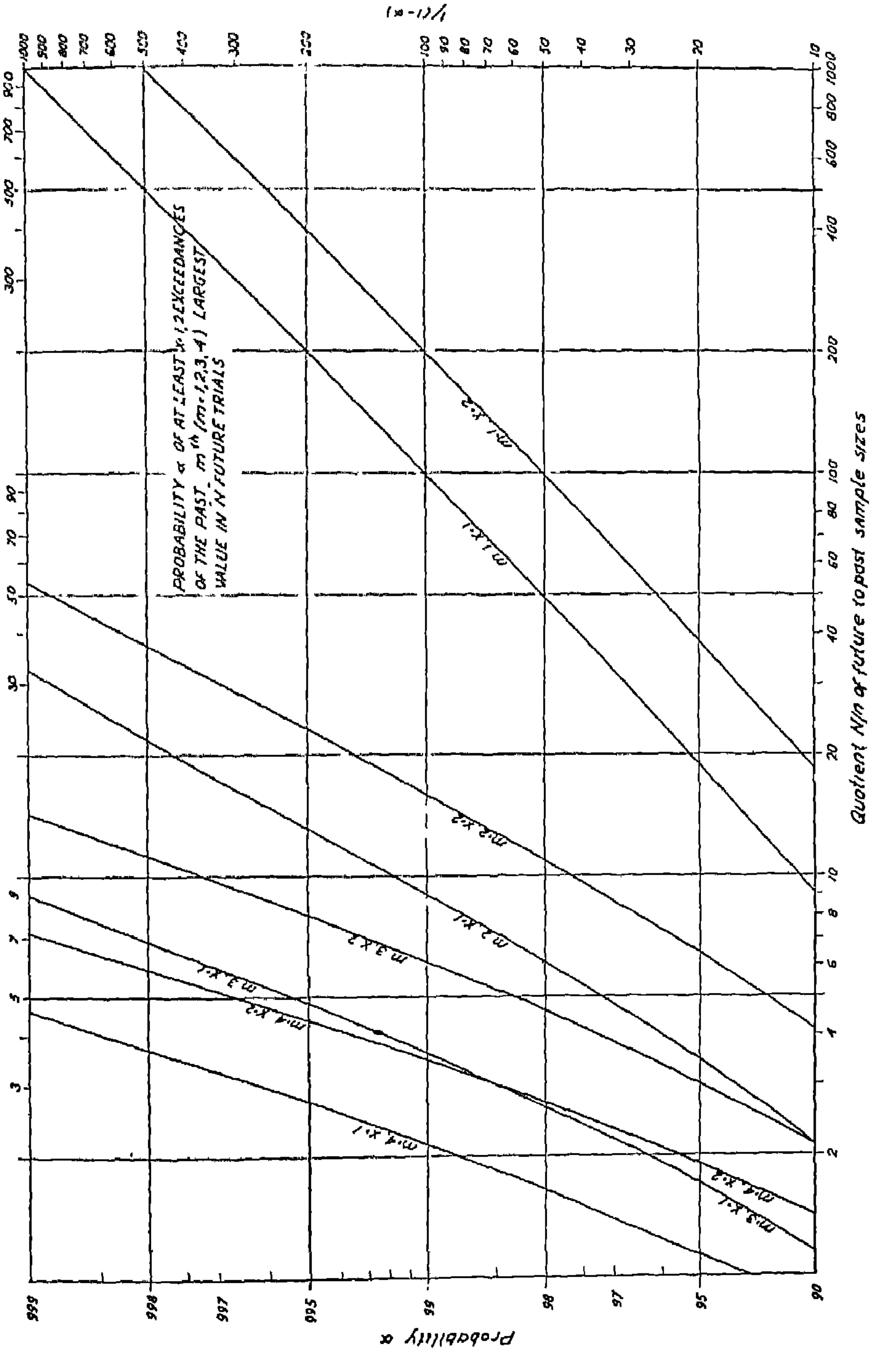
$$\frac{1}{1 - \alpha_2} = \frac{\left(\frac{N}{n} + 1\right)^2}{\frac{2N}{n} + 1}. \tag{5.7}$$

The probability α_2 as function of N/n is also traced in Graph (3) and designated by $m = 1, x = 2$. Finally, for $m = 2$ the probability α_2 for the penultimate value to be exceeded at least twice is obtained for large n by

$$\frac{1}{1 - \alpha_2} = \frac{\left(\frac{N}{n} + 1\right)^3}{\frac{3N}{n} + 1}. \tag{5.8}$$

This probability α_2 is also traced in Graph 3 and designated by $m = 2, x = 2$. If we fix the probabilities α_2, the graph shows the number of future trials corresponding to 1 and 2 exceedances over the largest, the penultimate, and the two preceding observations.

GRAPH 3


PROBABILITY α OF AT LEAST x=1,2 EXCEEDANCES OF THE PAST m^th (m=1,2,3,4) LARGEST VALUE IN N FUTURE TRIALS
Quotient N/n of future to past sample sizes
Probability α
1/(1-α)
m=1 x=1
m=1 x=2
m=2 x=1
m=2 x=2
m=3 x=1
m=3 x=2
m=4 x=1
m=4 x=2

6. Applications. In 50% of all cases, the largest (or smallest) of n past observations will not (or always) be exceeded in $N = n$ future trials. The mean number of exceedances is the mean in the Bernoulli distribution. *The variance is largest for the median, and smallest for the extremes*, and this superiority of the extremes increases with the sample size.

If the previous, and the future sample sizes both are large and equal, the distribution of the number of exceedances over the median observation is normal with mean and variance of the order $n/2$, whereas the distribution of the exceedances over the mth extremes (the law of rare exceedances), similar to the Poisson distribution, has the mean m, and the variance $2m$, m being small compared to the sample size. Elementary calculations lead to the setting of sample sizes N corresponding to given probabilities for 1 or 2 exceedances over the past largest and penultimate observation.

These methods may be of interest for forecasting floods if, instead of the size of the flood, we are interested only in the frequency. The same procedure may also be applied to other meteorological phenomena such as droughts, the extreme temperatures (the killing frost), the largest precipitations, etc., and permits to forecast the number of cases surpassing a given severity within the next N years.

REFERENCES

[1] E. J. GUMBEL, "Simple tests for given hypotheses," *Biometrika*, Vol. 32 (1942).

[2] H. VON SCHELLING, "A formula for the partial sums of some hypergeometric series," *Annals of Math. Stat.*, Vol. 20 (1949), No. 1.

[3] S. S. WILKS, "Statistical predictions with special reference to the problem of tolerance limits," *Annals of Math. Stat.*, Vol. 13 (1942).

ON THE ASYMPTOTIC DISTRIBUTION OF THE SUM OF POWERS OF UNIT FREQUENCY DIFFERENCES[1]

By Bradford F. Kimball

New York State Department of Public Service

1. Summary. Since the "unit" frequency differences (see (2.2) below) are dependent, the usual methods for establishing the normal character of the asymptotic distribution of the sum of random variables fail.

However, the essential character of the distribution is disclosed by the integral functional relationship (3.6). From this it is possible to show that for large samples the distribution approximates "stability" in the normal sense ([2] and Lemma 2).

Using the condition that the third logarithmic derivative of the characteristic function is uniformly bounded for all n on a neighborhood of $t = 0$ one can prove that the asymptotic distribution exists and is normal.

2. Introduction. Consider a one dimensional statistical universe characterized by a cumulative frequency function (*cdf*) $F(x)$ which is continuous. Consider an ordered random sample x_i of size N such that

$$x_i \leq x_{i+1}, \qquad i = 1 \text{ to } N - 1. \tag{2.1}$$

Consider frequency differences u_i defined by

$$\begin{aligned} u_1 &= F(x_1), \qquad u_{N+1} = 1 - F(x_N), \\ u_{i+1} &= F(x_{i+1}) - F(x_i), \qquad i = 1 \text{ to } N - 1. \end{aligned} \tag{2.2}$$

Thus

$$\sum_{N+1} u_i \equiv 1, \tag{2.3}$$

and the formal integral of the probability density function (*pdf*) of the u_i taken over the complete sample space of x_i can be written as

$$N! \int du_1\, du_2 \cdots du_{h-1}\, du_{h+1} \cdots du_{N+1} = 1, \tag{2.4}$$

where u_h is any u_i which it is found convenient to omit, and the region of integration is the N-fold Euclidean space bounded by the coordinate hyperplanes

$$u_i = 0, \qquad i \neq h, \qquad i = 1, 2, \cdots N + 1,$$

and the hyperplane

$$u_1 + u_2 + \cdots + u_{h-1} + u_{h+1} + \cdots + u_{N+1} = 1. \tag{2.5}$$

(See [1]).

[1] This is the second paper in connection with the subject announced in Abstract No. 9, *Annals of Math Stat.*, Vol 17 (1946), p 502, and Abstract No. 331, *Bull Am Math Soc*, Vol. 52 (1946), p 827 For first paper, see [1]

Consider a test function y_M defined by

$$y_M = \sum_M u_i^p, \qquad p > 0, \qquad M \leq N + 1, \tag{2.6}$$

where p is a real positive number, M is an integer less than or equal to $N + 1$ and such that if $M < N + 1$ the u_i which are to be omitted may be arbitrarily selected, but the subscripts indicating the order relation (2 2) are for the present retained.

Consider the case where N is odd and M is even, and set

$$N = 2n + 1, \qquad M = 2m. \tag{2.7}$$

Divide the set of $N + 1$ frequency differences u_i defined by (2 2) into two subsets such that each subset contains $n + 1$ differences of which exactly m are included in the test function (2.6). Now let N become infinite over odd numbers $N_1, N_2, \cdots$. In other words the sample size is to increase without limit. For each sample size N_j in such a sequence let M_j be an even number such that

$$M_j \leq N_j + 1 \tag{2 8}$$

and such that the ratio M_j/N_j is controlled for large values of N by

$$\lim_{h \to \infty} M_j/N_j = \text{constant } c, \qquad 0 < c \leq 1. \tag{2.9}$$

As above for each step in the sequence the set of $N_j + 1$ frequency differences u_i is divided into two subsets of $n_j + 1$ frequencies each with

$$N_j = 2n_j + 1, \qquad M_j = 2m_j, \tag{2.10}$$

such that m_j frequencies of each subset are included in the test function

$$y_{M_j} = \sum u_i^p. \tag{2.11}$$

Now we note that for a random sample of size N taken from the above universe, the characteristic function $G_N(t; y_M)$ may be defined by

$$G_N(t; y_M) = N! \int e^{ity_M}\, du_1\, du_2 \cdots du_N \tag{2 12}$$

taken over region in Euclidean space of N dimensions as indicated for the integral (2.4), taking index h equal to $N + 1$.

3. Proof of integral relationship—Lemma 1. For simplicity of notation drop subscripts from M_j, N_j, n_j and m_j. We separate the test function y_M into two parts y_m and $y_{m'}$ such that

$$y_M = y_m + y_{m'} = \sum_m u_i^p + \sum_{m'} u_i^p, \qquad m = m' = M/2 \tag{3 1}$$

where the m frequency differences u_i in y_m are those included in first subset and those contained in $y_{m'}$ are those of the original M frequencies included in the second subset (see (2.10) and (2.11))

The formal integral defining $G_N(t; y_M)$ may be written

$$(3.2) \quad G_N(t, y_M) = \Gamma(2n+2) \int_{R_2} e^{ity_m} \, du_1 \cdots du_{n+1} \int_{R_1} e^{ity_{m'}} \, du_{n+2} \cdots du_{2n+1},$$

where

R_2 = $2n+1$ dimensional Euclidean space bounded by coordinate hyperplanes and plane $\sum_{2n+1} u_i = 1$,

R_1 = n dimensional Euclidean space bounded by the coordinate hyperplanes and the plane

$$(3.3) \quad \begin{aligned} u_{n+2} + u_{n+3} + \cdots + u_{2n+1} &= 1 - w, \\ w &= u_1 + u_2 + \cdots + u_{n+1}. \end{aligned}$$

Now introduce the transformation to u_i'

$$(3.4) \quad u_i'(1-w) = u_i, \qquad i = n+2, n+3, \cdots, 2n+1, 2n+2.$$

Thus we have

$$\sum_{n+1} u_i' \equiv 1,$$

and the n u_i' involved in the integration are bounded above by the hyperplane $\sum_n u' = 1$. The Jacobian is $(1-w)^n$.

Similarly under transformation

$$(3.5) \quad \begin{aligned} v_i w = u_i, \qquad i &= 1, 2, \cdots, n+1, \\ \sum_{n+1} v_i &\equiv 1 \end{aligned}$$

Let v_i, $i = 1, 2, \cdots n$ and w replace the remaining variables of integration. Thus the region of integration of these v_i is $v_i \geq 0$ with the hyperplane $\sum_n v_i = 1$ furnishing the upper bound. The Jacobian of the transformation is w^n.

The regions of integration of these new variables u_i' and v_i are seen to be independent of each other and of w. Noting effect of above transformations on y_m and $y_{m'}$, the integral (3.2) will be found to reduce to the following form:

$$(3.6) \quad G_N(t; y_M) = \frac{\Gamma(2n+2)}{\Gamma^2(n+1)} \int_0^1 w^n (1-w)^n G_n(tw^p; y_m) G_n(t(1-w)^p; y_m) \, dw,$$

where

$$N = 2n+1, \qquad M = 2m.$$

Lemma 1. *This functional relationship holds for all values of N and M subject to the condition that N be an odd integer and M an even integer. One may note that a similar integral functional relationship will hold for any partition $(n_0 n_1)$ of the $N-1$ free frequency differences such that*

$$n_0 + n_1 = N - 1, \qquad m_0 + m_1 = M,$$

with corresponding changes in the Gamma functions which precede the integral.

In order to find out what happens when N becomes large the partially normalized test function z_M is introduced. This is defined by

$$z_M = (y_M - \bar{y}_M)(N + 1)^p/\sqrt{M}, \tag{3.7}$$

where (cf. [1], formula (3.1))

$$\bar{y}_M = E(y_M) = \frac{M\Gamma(N + 1)\Gamma(p + 1)}{\Gamma(N + 1 + p)}. \tag{3.8}$$

I have referred to z_M as a partially normalized variable since

$$\begin{aligned} &E(z_M) = 0, \\ &\lim_{N\to\infty} E(z_M^2) = \Gamma(2p + 1) - \Gamma^2(p + 1) - cp^2\Gamma^2(p + 1), \end{aligned} \tag{3.9}$$

where this limit can be shown to be greater than zero for

$$\begin{aligned} p &\neq 1, \qquad 0 < c \leq 1, \\ p &= 1, \qquad 0 < c < 1. \end{aligned} \tag{3.10}$$

Recalling the separation of the test function into two parts (see (3.1)) we define $\bar{y}_m$ and $\bar{y}_{m'}$ by

$$\bar{y}_m = \bar{y}_{m'} = \frac{m\Gamma(n + 1)\Gamma(p + 1)}{\Gamma(n + 1 + p)} \tag{3.11}$$

with

$$M = 2m, \qquad N = 2n + 1.$$

From Stirling's formula it can then be shown that

$$(N + 1)^p\bar{y}_M/\sqrt{M} = (2^p/\sqrt{2})2[(n + 1)^p\bar{y}_m/\sqrt{m}] + o(1), \tag{3.12}$$

where $o(1)$ goes to zero as N and M become infinite subject to the condition (2.9). Thus if we define z_m and $z_{m'}$ by

$$z_m = (y_m - \bar{y}_m)(n + 1)^p/\sqrt{m}, \qquad z_{m'} = (y_{m'} - \bar{y}_{m'})(n + 1)^p/\sqrt{m}, \tag{3.13}$$

since

$$y_M = y_m + y_{m'}$$

and

$$(N + 1)^p/\sqrt{M} = (2^p/\sqrt{2})(n + 1)^p/\sqrt{m},$$

it follows that

$$z_M = (2^p/\sqrt{2})(z_m + z_{m'}) + o(1). \tag{3.14}$$

Hence if we denote the characteristic function of the distribution of the

partially normalized test function z_M by $G_N(t; z_M)$ and proceed to develop an integral functional relationship similar to (3.6), one arrives at

$$G_N(t; z_M) = e^{ito(1)} \frac{\Gamma(2n+2)}{\Gamma^2(n+1)} \int_0^1 w^n(1-w)^n G_n[t(2w)^p/\sqrt{2}, z_m] \cdot G_n[t2^p(1-w)^p/\sqrt{2}; z_m]\, dw \tag{3.15}$$

with

$$N = 2n + 1, \qquad M = 2m$$

4. Resulting functional relationship when N becomes large. The second lemma shows that the functional equation satisfied by the characteristic function of a normal distribution is approximated when N is large. Suppose we now set

$$w = (1 + s)/2, \qquad 1 - w = (1 - s)/2, \qquad dw = ds/2. \tag{4.1}$$

Substituting in (3.15) we have

$$G_N = \frac{e^{ito(1)}\,\Gamma(2n+2)}{2^{2n+1}\,\Gamma^2(n+1)} \int_{-1}^{+1} (1 - s^2)^n G_n[t(1+s)^p/\sqrt{2}; z_m] \; G_n[t(1-s)^p/\sqrt{2}, z_m]. \tag{4.2}$$

Set

$$H(t, s) = G_n[t(1+s)^p/\sqrt{2}; z_m] G_n[t(1-s)^p/\sqrt{2}; z_m]. \tag{4.3}$$

Then

$$H_s = G_n' G_n tp(1+s)^{p-1}/\sqrt{2} - G_n G_n' tp(1-s)^{p-1}/\sqrt{2}. \tag{4.4}$$

Using law of mean write

$$H(t, s) = H(t, 0) + sH_s[t, h(s)], \qquad 0 < |h(s)| < s. \tag{4.5}$$

Substituting in (4.2) we have

$$e^{-ito(1)} G_N = H(t, 0) + \frac{\Gamma(2n+2)}{2^{2n}\,\Gamma^2(n+1)} \int_0^1 H_s[t, h(s)](1 - s^2)^n s\, ds \tag{4.6}$$

With $E(z_m) \equiv 0$, from the fact that the limiting variance of z_m is bounded (see (3.9)) it follows that the first derivative of its characteristic function remains bounded in any finite interval, for all n ([3], p 90). Thus

$$|G_n'(t; z_m)| < A, \qquad 0 \le |t| \le D, \qquad \text{for all } n \tag{4.7}$$

For case $p \geq 1$, by virtue of condition (4.7) H_s will remain bounded over interval of integration of (4.6) as N becomes infinite. Let B denote such upper bound of the absolute value of H_s. Then, carrying out the integration

$$\text{absolute value of integral} < \frac{B\Gamma(2n+2)}{2^{2n}\,\Gamma^2(n+1)} \frac{1}{2(n+1)} \tag{4.8}$$

for any value of t. This quantity approaches zero as N goes to infinity uniformly for t on any finite range. For the case that $0 < p < 1$ a similar argument may be used by including the factor $(1 - s)^{p-1}$ which appears in H_n in the integration, and placing the upper bound on the absolute value of the factor $G_n G_n'$.

Substituting back for $H(t, 0)$ in (4.6) one arrives at

LEMMA 2. *The characteristic function $G_n(t; z_m)$ satisfies the relationship*

$$G_N(t; z_M) = [G_n(t/\sqrt{2}; z_m)]^2 + o(1), \qquad N = 2n + 1, \qquad M = 2m, \tag{4.9}$$

where $o(1)$ goes to zero with increasing n, uniformly for t on any finite interval

$$0 \leq |t| \leq D. \tag{4.10}$$

The above lemma indicates that if the asymptotic *pdf* of z_m exists, it will be a "stable" distribution in the normal sense [2]. In order to set the stage for proving the existence of this asymptotic distribution we shall first investigate the third logarithmic derivative of $G_n(t; z_m)$.

5. Investigation of third logarithmic derivative. We shall now show that the third logarithmic derivative of G is uniformly bounded in some neighborhood of $t = 0$. We first prove that the absolute value of the third derivative of G is bounded for all t and n. Now the third derivative will have absolute value less than the third absolute moment which I denote by μ_3. Using Liapounoff's inequality

$$\mu_3^2 \leq \mu_2 \mu_4 \tag{5.1}$$

one asks whether the fourth moment μ_4 remains finite as n and m become infinite.

Computation of the fourth moment about the mean appears to be somewhat formidable. However it is not so difficult to show that it remains finite with increasing m and n. Referring to previous paper ([1] formulas (4.8)–(4.10)) we use quasi-moment generating function $g_0(x)$ such that

$$d^r g_0(0)/dx^r = \Gamma(pr + 1), \qquad g_0(0) = 1, \tag{5.2}$$

and it follows that

$$E(\sum_m u_i^p)^r = d^r[g_0(0)]^m/dx^r \Gamma(n + 1)/\Gamma(n + 1 + pr), \tag{5.3}$$

and one recalls that

$$y = \sum_m u_i^p, \qquad \bar{y} = \frac{m\Gamma(n + 1)\Gamma(p + 1)}{\Gamma(n + 1 + p)}$$

with

$$z = [(n + 1)^p/\sqrt{m}][y - \bar{y}].$$

The resulting fourth moment of z will be in the form of a fourth degree polynomial in m whose coefficients are of the type

$$\frac{(n + 1)^{4p}\Gamma(n + 1)}{\Gamma(n + 1 + 4p)}, \qquad \frac{(n + 1)^{3p}\Gamma(n + 1)}{\Gamma(n + 1 + 3p)}, \cdots,$$

combined with the first moment, with m^{-2} appearing as a factor. By expansion of the Gamma function in asymptotic series in $(n + 1)$ it is not difficult to show that the coefficient of m^{1} becomes asymptotic like $(n + 1)^{-2}$, and that the coefficient of m^{0} becomes asymptotic like $(n + 1)^{-1}$. It follows that as n and m go to infinity with $m \sim c(n + 1)$, that this fourth moment approaches a finite limit. Hence one concludes that the third derivative of G has bounded absolute value for all n and t

Since the absolute value of the first derivative of G is uniformly bounded for finite t and all n it follows from the properties of a characteristic function that given a positive number K less than unity, it is possible to find a value of $t = t_0$ greater than zero such that

$$(5.4) \qquad 0 < K \leq | G_n(t, z) | \leq 1, \qquad 0 \leq | t | \leq t_0 ,$$

for all n.

From the above double inequality and the fact that the absolute values of the first three derivatives are uniformly bounded it follows that *the third logarithmic derivative of G is uniformly bounded for all n on the interval*

$$(5.5) \qquad 0 \leq | t | \leq t_0 .$$

6. Proof that the asymptotic distribution of z exists and is normal. Since absolute value of G is uniformly bounded away from zero on interval (5.5) one can write the functional relation (4.9) as

$$(6.1) \qquad \log G_N(t, z_M) = 2 \log G_n(t/\sqrt{2}, z_m) + o(1),$$

where $o(1)$ goes to zero with increasing n uniformly for t on interval (5.5).

Introduce the notation:

$\lambda(n)$ equals variance of z_m ,

$q(t, n)$ equals third logarithmic derivative of $G_n(t, z_m)$,

$R(t, N)$ equals remainder defined by

$$(6.2) \qquad \log G_N(t, z_M) = -\lambda(N)t^2/2 + R(t, N).$$

Write

$$(6.3) \quad \log G_n(t/\sqrt{2}, z_m) = -\lambda(n)t^2/4 + q(t\theta/\sqrt{2}, n)t^3/(12\sqrt{2}), \quad 0 < \theta < 1.$$

Substituting (6.2) and (6.3) in (6.1)

$$(6.4) \quad R(t, N) = [\lambda(N) - \lambda(n)]t^2/2 + [1/\sqrt{2}]q(t\theta/\sqrt{2}, n)t^3/6 + o(1).$$

By (3.9)

$$(6.5) \qquad \lim \lambda(n) = \lim \lambda(N) = \text{positive number } \lambda.$$

We have proved that there exists an upper bound U such that

$$(6.6) \qquad | q(t, n) | \leq U$$

for all n and for t on interval

$$(6.7) \qquad 0 \leq | t | \leq t_0 .$$

Hence from (6.4) one can reason that given a positive ϵ, a number N_0 can be found such that

$$(6.8) \qquad | R(t, N) | \leq [1/\sqrt{2}]U | t^3/6 | + \epsilon$$

for all t on (6.7) and for $N > N_0$.
By (6.1)

$$(6.9) \quad R(t, 2N + 1) = [\lambda(2N + 1) - \lambda(N)]t^2/2 + 2R(t/\sqrt{2}, N) + o(1).$$

Using (6.8)

$$| R(t/\sqrt{2}, N) | \leq [1/\sqrt{2}]U | t^3/(12\sqrt{2}) | + \epsilon.$$

Hence for any positive number ϵ_2 a number N_2 can be found such that

$$| R(t, N) | \leq (1/2)U | t^3/6 | + 2\epsilon + \epsilon_2 , \qquad N > N_2 ,$$

for all t on (6.7). After k such operations, taking $\epsilon_s = \epsilon$

$$(6.10) \qquad | R(t, N) | \leq (1/2)^{k/2}U | t^3/6 | + (2^k - 1)\epsilon, \qquad N > N_k .$$

Thus given a positive number d one can determine k such that

$$(1/2)^{k/2}Ut_0^3/6 < d/2,$$

and ϵ such that

$$2^k\epsilon < d/2,$$

and therefore a number N_{k+1} such that

$$(6.11) \qquad | R(t, N) | < d, \qquad N > N_{k+1}$$

for all t on interval (6.7).

It follows that $G_N(t, z_M)$ converges uniformly to exp $(-\lambda t^2/2)$ on interval (6.7).

Convergence of $G_N(t, z_M)$ for a value $t = t_1$ outside the interval (6.7) may be proved by choosing integer k such that

$$(6.12) \qquad 0 < | t_1 |/(\sqrt{2})^k \leq t_0 ,$$

and taking

$$t_3 = t_1/(\sqrt{2})^k.$$

Recalling that the functional relation (4.9) holds for all finite t, this can be applied k times, thus building up t_3 to t_1

It follows from the continuity theorem that the distribution function of z_m converges to the normal distribution function.

7. Statement of theorem proved. The proof given above has involved the restriction that N be odd and M even (see (2 7)). This restriction is required

for the integral relationship (3.6). However, if N were even one could take $n_0 = N/2$ and $n_1 = n_0 - 1$ and deal with, G_{n_0} and G_{n_1} in the integrand. Also if M were odd, one could take $m_0 = (M + 1)/2$, $m_1 = m_0 - 1$, and deal with $G_{n_0}(t, m_0)$ and $G_{n_1}(t, m_1)$ in the integrand. This would of course carry with it corresponding changes in the Gamma functions which precede the integral. As long as we require that

$$N = n_0 + n_1 + 1, \qquad M = m_0 + m_1,$$
$$\lim M/N = \lim m_0/n_0 = \lim m_1/n_1 = c > 0,$$

the arguments used in arriving at the asymptotic relations (3.15) and (4.9) will apply. Hence the theorem:

THEOREM[2]. *For a one dimensional statistical universe whose cdf is continuous, consider the function of the unit frequency differences u,*

$$y = \sum_m u_i^p \tag{7.1}$$

taken from an ordered random sample of size n (see (2.2)) where p is any real positive number, and m is any positive integer less than or equal to $n + 1$. The selection of which m unit frequencies are to be included is arbitrary. Then with

$$\bar{y} = E(y) = \frac{m\Gamma(n + 1)\Gamma(p + 1)}{\Gamma(n + p + 1)} \tag{7.2}$$

consider the partially normalized variable

$$z = \frac{(n + 1)^p}{\sqrt{m}} (y - \bar{y}). \tag{7.3}$$

If n goes to infinity, with m becoming infinite so that

$$\lim m/n = c > 0, \tag{7.4}$$

then the asymptotic cumulative distribution of z exists and is normal, with

$$\lim E(z^2) = \Gamma(2p + 1) - \Gamma^2(p + 1) - cp^2\Gamma^2(p + 1), \tag{7.5}$$

except in the trivial case $p = 1$, $m = n + 1$, in which case $z \equiv 0$, and in the case $p = 1$, $c = 1$.

REFERENCES

[1] B. F. KIMBALL, "Some basic theorems for developing tests of fit for the case of the non-parametric probability distribution, I", *Annals of Math. Stat.*, Vol. 18 (1947), No. 4, pp. 540-548

[2] P. LEVY, *Théorie de l'Addition des Variables Aléatoires*, Gauthier-Villars, Paris, 1937, Chapter V.

[3] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[4] P. A. P. MORAN, "Random Division of an Interval", *Jour. Roy. Stat. Assoc.*, *Suppl.*, Vol. 9 (1947), pp. 92-98

[2] For the case $p = 2$, $m = n + 1$, an interesting proof was published by P. A. P. Moran in 1947, see [4]

EFFECT OF LINEAR TRUNCATION ON A MULTINORMAL POPULATION[1]

By Z. W. Birnbaum[2]

University of Washington

1. Introduction. In admission to educational institutions, personnel selection, testing of materials, and other practical situations, the following mathematical model is frequently encountered: A $(k+l)$-dimensional random variable $(X_1, X_2, \cdots, X_k, Y_1, Y_2, \cdots, Y_l) = (\mathbf{X}, \mathbf{Y})$ is considered, with a joint probability-distribution assumed to be non-singular multi-normal. The $Y_1, Y_2, \cdots, Y_l$ are scores in admission tests, the $X_1, X_2, \cdots, X_k$ scores in achievement tests. The admission tests are administered to all individuals in the $(\mathbf{X}, \mathbf{Y})$ population to decide on admission or rejection, and (usually at some later time) the achievement tests are administered to those admitted. A set of weights $a_j \geq 0$, $j = 1, 2, \cdots, l$ is used to define a composite admission test score $U = \sum_{j=1}^{l} a_j Y_j$ and a "cutting score" τ is chosen so that an individual is admitted if $U \geq \tau$, and rejected if $U < \tau$. We will refer to this procedure as *linear truncation of* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathbf{Y}$ *to the set* $U \geq \tau$.

A linear truncation in $\mathbf{Y}$ clearly will change the absolute distribution of $\mathbf{X}$, except in the case of independence. In this paper a study is made of the absolute distribution of $\mathbf{X}$ after linear truncation in $\mathbf{Y}$ in the case $k = 1$; in particular, the possibility is investigated of choosing the a_j and τ in such a way that the distribution of $\mathbf{X}$ after truncation has certain desirable properties. The case $k > 1$ leads to a considerable diversity of problems which are being studied and, it is hoped, will be the subject of a separate paper.

Throughout this paper it will be assumed that all the parameters of $(\mathbf{X}, \mathbf{Y})$, that is the expectations, variances and covariances before truncation, are known. In practical situations it often happens that only the parameters of $Y_1, Y_2, \cdots, Y_l$ before truncation are known, while the first and second moments involving $X_1, X_2, \cdots, X_k$ are only known for the joint distribution after truncation. It can be shown [1] that in such situations the expectations, variances and covariances of $(\mathbf{X}, \mathbf{Y})$ before truncation can always be reconstructed if $(\mathbf{X}, \mathbf{Y})$ has a multinormal distribution.

In the simplest case $k = l = 1$ the probability-density of the original binormal random variable (X, Y) may be, without loss of generality, assumed equal to

$$f(X, Y, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-(X^2 - 2\rho XY + Y^2)/2(1-\rho^2)}. \tag{1.1}$$

By truncating this distribution in Y to the set $Y \geq \tau$ one obtains the probability-density

$$g(X, Y; \rho, \tau) = \begin{cases} \psi^{-1}(\tau) f(X, Y; \rho), & \text{for } Y \geq \tau, \\ 0, & \text{for } Y < \tau, \end{cases} \tag{1.2}$$

[1] Presented to the Institute of Mathematical Statistics on June 18, 1949.

[2] Research done under the sponsorship of the Office of Naval Research.

where

(1.3) $$\psi(\tau) = \frac{1}{\sqrt{2\pi}} \int_\tau^\infty e^{-t^2/2}\, dt.$$

For further use we introduce the abbreviations

(1.4) $$\varphi(\tau) = \frac{1}{\sqrt{2\pi}} e^{-\tau^2/2},$$

(1.5) $$\lambda(\tau) = \frac{\varphi(\tau)}{\psi(\tau)}.$$

We also note the inequalities

(1.6) $$\tau \leq \lambda(\tau)$$

and

(1.7) $$\lambda(\tau) \leq \frac{\sqrt{4 + \tau^2} - \tau}{2}$$

derived in [2] and [3], respectively.[3]

Before proceeding to the more-dimensional case, we will study some properties of the marginal probability-distribution of X after truncation to $Y \geq \tau$

(1.8) $$\varphi(X; \rho, \tau) = \int_\tau^\infty g(X, Y; \rho, \tau)\, dY.$$

2. The moments of $\varphi(X; \rho, \tau)$. In this section all mathematical expectations are computed for the absolute distribution of X after truncation of (X, Y) to $Y \geq \tau$.

We have

$$\varphi(X; \rho, \tau) = \psi^{-1}(\tau)\varphi(X)\psi\left(\frac{\tau - \rho X}{\sqrt{1 - \rho^2}}\right),$$

and hence

$$\begin{aligned} E(X^n) &= \int_{-\infty}^{+\infty} X^n \varphi(X, \rho, \tau)\, dX \\ &= \psi^{-1}(\tau) \int_{-\infty}^{+\infty} \frac{X}{\sqrt{2\pi}} e^{-X^2/2} \frac{X^{n-1}}{\sqrt{2\pi}} \int_{(\tau-\rho X)/\sqrt{1-\rho^2}}^\infty e^{-S^2/2}\, dS\, dX \\ &= \psi^{-1}(\tau) \left\{ -\varphi(X)X^{n-1}\psi\left(\frac{\tau - \rho X}{\sqrt{1 - \rho^2}}\right) \Big|_{x=-\infty}^{+\infty} \right. \\ &\quad + \int_{-\infty}^{+\infty} \varphi(X) \left[\frac{dX^{n-1}}{dX}\, \psi\left(\frac{\tau - \rho X}{\sqrt{1 - \rho^2}}\right)\right. \end{aligned}$$

[3] Implicitly, the inequality (1.6) was known already to Laplace, cf. *Mécanique Céleste*, transl. by Bowditch, Boston 1839, Vol. 4, p. 493.

$$+ X^{n-1} \frac{\rho}{\sqrt{1-\rho^2}} \varphi\left(\frac{\tau - \rho X}{\sqrt{1-\rho^2}}\right)\Bigg] dX \Bigg\}$$

$$= E\left(\frac{dX^{n-1}}{dX}\right) + \frac{\rho}{\psi(\tau)\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} X^{n-1}\varphi(X)\varphi\left(\frac{\tau - \rho X}{\sqrt{1-\rho^2}}\right) dX.$$

From the identity

$$\varphi(X)\varphi\left(\frac{\tau-\rho X}{\sqrt{1-\rho^2}}\right) = \varphi(\tau)\varphi\left(\frac{X-\rho\tau}{\sqrt{1-\rho^2}}\right) \tag{2.0}$$

we obtain

$$\int_{-\infty}^{+\infty} X^{n-1}\varphi(X)\varphi\left(\frac{\tau-\rho X}{\sqrt{1-\rho^2}}\right) dX = \varphi(\tau)\int_{-\infty}^{+\infty} X^{n-1}\varphi\left(\frac{X-\rho\tau}{\sqrt{1-\rho^2}}\right) dX$$

$$= \sqrt{1-\rho^2}\,\varphi(\tau)\int_{-\infty}^{+\infty} (S\sqrt{1-\rho^2}+\rho\tau)^{n-1}\varphi(S)\, dS,$$

and hence

$$E(X^n) = E\left(\frac{dX^{n-1}}{dX}\right) + \rho\lambda(\tau)\int_{-\infty}^{+\infty} (S\sqrt{1-\rho^2}+\rho\tau)^{n-1}\varphi(S)\, dS, \tag{2.1}$$

for $n \geq 1$.

For $n = 1$ this yields for the expectation of X after truncation

$$E(X) = \rho\lambda(\tau). \tag{2.2}$$

For $n = 2$ we have from (2.1)

$$E(X^2) = 1 + \rho^2\tau\lambda(\tau) = 1 + \rho\tau E(X),$$

and hence for the variance of X after truncation the expression

$$\sigma^2(X) = 1 + E(X)[\rho\tau - E(X)], \tag{2.3}$$

or

$$\sigma^2(X) = 1 - \rho^2\lambda(\tau)[\lambda(\tau) - \tau]. \tag{2.31}$$

From (2.2) we see that $E(X)$ always has the sign of ρ, as one would expect. From (2.3) one finds a lower bound for τ

$$\tau > \frac{E^2(X) - 1}{\rho E(X)}. \tag{2.4}$$

From (2.31) and (1.6) one concludes that $\sigma^2(X) < 1$ for $\rho \neq 0$, hence the variance of X after truncation is always less than the variance before truncation, except if $\rho = 0$.

Similarly one computes from (2.1) the third moment about zero

$$E(X^3) = E(X)[3 - \rho^2(1 - \tau^2)]$$

and obtains for the third moment about the expectation

$$E[X - E(X)]^3 = E(X)\rho^2\{[\lambda(\tau) - \tau][2\lambda(\tau) - \tau] - 1\}. \tag{2.5}$$

Numerical computation indicates that the quantity in braces is always >0, which would mean that the skewness of X after truncation has the same sign as $E(X)$ and ρ. No analytic proof of this statement has been obtained.

3. Determination of τ for given expectation or quantile of X after truncation; dependence of this τ on ρ. Let it be required to determine τ so that the expectation of X after truncation assumes a given value m. It follows immediately from (2.2) that this τ is obtained by solving the equation

$$\lambda(\tau) = \frac{m}{\rho} \tag{3.1}$$

for τ, which can be done with the aid of a table[4] of $\lambda(\tau)$.

Another problem which occurs in applications consists in determining τ so that, for given $0 < \alpha < 1$ and X_α, the α-quantile for X after truncation assumes the value X_α, that is so that

$$\int_{-\infty}^{X_\alpha} \varphi(X; \rho, \tau)\, dX = \psi^{-1}(\tau) \int_{-\infty}^{X_\alpha} \int_{\tau}^{\infty} f(X, Y; \rho)\, dY\, dX = \alpha. \tag{3.2}$$

Let

$$P(s, t; \rho) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \int_s^\infty \int_t^\infty e^{-[(X^2 - 2\rho XY + Y^2)/2(1-\rho^2)]}\, dY\, dX \tag{3.21}$$

denote the volume of the probability solid $Z = f(X, Y; \rho)$ above the quadrant $X \geq s$, $Y \geq t$. Then (3.2) may be written in the form

$$1 - \frac{P(X_\alpha, \tau, \rho)}{\psi(\tau)} = \alpha,$$

or

$$(1 - \alpha)\psi(\tau) = P(X_\alpha, \tau; \rho), \tag{3.3}$$

and this equation can be solved for τ by trial with the aid of tables of $\psi(\tau)$ and Pearson's tables [4] of $P(s, t; \rho)$.

LEMMA 1. *For fixed expectation of X after truncation $E(X) = m$, the solution $\tau(\rho)$ of (3.1) is a strictly decreasing function of the absolute value of ρ for $0 < |\rho| \leq 1$.*

PROOF: Differentiating $m = \rho\lambda(\tau)$ with regard to ρ one obtains

$$0 = \lambda(\tau) + \rho\lambda'(\tau)\frac{d\tau}{d\rho}$$

and, in view of the identity

$$\lambda'(\tau) = \lambda(\tau)[\lambda(\tau) - \tau],$$

the expression

$$\frac{d\tau}{d\rho} = -\frac{1}{\rho[\lambda(\tau) - \tau]}. \tag{3.4}$$

---

[4] A table of $1/\lambda(\tau)$ is, for example, given in Karl Pearson, *Tables for Statisticians and Biometricians, Part II*, 1931, pp. 11–15.

From (3.4) and (1.6) we see that

$$\operatorname{sign} \frac{d\tau}{d\rho} = -\operatorname{sign} \rho,$$

which proves our lemma.

LEMMA 2. *For fixed α, X_α, the solution $\tau = \tau(\rho)$ of (3.3) is a strictly decreasing function of $|\rho|$ for $0 < |\rho| \leq 1$.*

PROOF: Differentiating (3.3) with regard to ρ one obtains

$$-(1 - \alpha)\varphi(\tau) \frac{d\tau}{d\rho} = \frac{\partial P}{\partial \tau} \frac{d\tau}{d\rho} + \frac{\partial P}{\partial \rho},$$

and hence

$$\frac{d\tau}{d\rho} = \frac{-\dfrac{\partial P}{\partial \rho}}{\dfrac{\partial P}{\partial \tau} + (1 - \alpha)\varphi(\tau)}. \tag{3.6}$$

From (3.21) one easily verifies that

$$\frac{\partial P(X_\alpha, \tau, \rho)}{\partial \rho} = \varphi(\tau)(1 - \rho^2)^{-\frac{1}{2}} \int_{(X_\alpha - \rho\tau)/\sqrt{1-\rho^2}}^{\infty} te^{-t^2/2}\, dt,$$

and therefore

$$\frac{\partial P(X_\alpha, \tau, \rho)}{\partial \rho} > 0. \tag{3.7}$$

One also computes

$$\frac{\partial P(X_\alpha, \tau; \rho)}{\partial \tau} = -\varphi(\tau)\psi\left(\frac{X_\alpha - \rho\tau}{\sqrt{1 - \rho^2}}\right),$$

so that the denominator of the right hand expression in (3.6) becomes

$$\varphi(\tau)\left[1 - \alpha - \psi\left(\frac{X_\alpha - \rho\tau}{\sqrt{1 - \rho^2}}\right)\right].$$

In view of (3.3) this is equal to

$$\begin{aligned}
\varphi(\tau) &\left[\frac{P(X_\alpha, \tau, \rho)}{\psi(\tau)} - \psi\left(\frac{X_\alpha - \rho\tau}{\sqrt{1 - \rho^2}}\right)\right] \\
&= \lambda(\tau)\left[P(X_\alpha, \tau; \rho) - \psi(\tau)\psi\left(\frac{X_\alpha - \rho\tau}{\sqrt{1 - \rho^2}}\right)\right] \\
&= \lambda(\tau) \frac{1}{2\pi} \int_\tau^\infty e^{-Y^2/2} \int_{(X_\alpha - \rho Y)/\sqrt{1-\rho^2}}^{(X_\alpha - \rho\tau)/\sqrt{1-\rho^2}} e^{-U^2/2}\, dU\, dY \\
&= \lambda(\tau) \frac{1}{2\pi} \int_\tau^\infty h(Y)\, dY.
\end{aligned}$$

If $\rho > 0$, then $\rho Y > \rho\tau$ in the interval of integration $\tau < Y < \infty$, hence $\frac{X_\alpha - \rho Y}{\sqrt{1-\rho^2}} < \frac{X_\alpha - \rho\tau}{\sqrt{1-\rho^2}}$, therefore the integrand $h(Y)$ is positive, and so is the denominator of (3.6). Similarly one sees that if $\rho < 0$ the integrand $h(Y)$ is negative for $\tau < Y < \infty$ and the denominator of (3.6) is negative. In view of (3.7) we conclude

$$\operatorname{sign} \frac{d\tau}{d\rho} = -\operatorname{sign} \rho \qquad \text{for } \rho \neq 0.$$

4. Linear truncation of $(X, Y_1, Y_2, \cdots, Y_l)$ to the set $\sum_{j=1}^{l} a_j Y_j \geq \tau$ for given expectation or quantile of X, minimizing the rejected part of the population. Let $(X, Y_1, Y_2, \cdots, Y_l)$ be an $(l+1)$-dimensional non-singular normal random variable with all expectations, variances and covariances known. We wish to choose $a_1, a_2, \cdots, a_l$ and τ so that by setting

$$U = \sum_{j=1}^{l} a_j Y_j \tag{4.1}$$

and performing the linear truncation to the set $U \geq \tau$ we obtain for the expectation of X after truncation a pre-assigned value m, and that this is achieved with the least waste of the original population, that is so that for the non-truncated probability-distribution the probability $P(\sum_{j=1}^{l} a_j Y_j < \tau)$ is minimum.

Without loss of generality we may assume that, before truncation, we have

$$E(X) = E(Y_1) = \cdots = E(Y_l) = 0, \tag{4.21}$$

$$\sigma^2(X) = 1, \tag{4.22}$$

and thus

$$E(U) = 0. \tag{4.3}$$

Furthermore, the a_j and τ can always be multiplied by a constant, without changing the set of truncation, so that we have

$$\sigma^2(U) = 1. \tag{4.4}$$

THEOREM 1. *To truncate $(X, Y_1, Y_2, \cdots, Y_l)$ linearly in $Y_1, Y_2, \cdots, Y_l$ so that the expectation of X after truncation has the given value m and that the probability of the rejected part of the original population is minimum, it is necessary and sufficient* (1) *to determine $a_1, a_2, \cdots, a_l$ so that the absolute value of the correlation-coefficient $\rho(X, U)$ becomes maximum under the condition* (4.4), *and* (2) *for U determined by these $a_1, a_2, \cdots, a_l$ and for $\rho = \rho(X, U)$ to solve equation* (3.1) *for τ.*

The proof of this theorem follows immediately from the first paragraph of section 3 and Lemma 1.

Using the second paragraph of section 3 and Lemma 2, one equally easily arrives at the following theorem:

THEOREM 2. *To truncate $(X, Y_1, Y_2, \cdots, Y_l)$ linearly in $Y_1, Y_2, \cdots, Y_l$*

so that the α-quantile of X after truncation has the given value X_α and that the probability of the rejected part of the original population is minimum, it is necessary and sufficient to satisfy (1) *in Theorem* 1 *and then to solve equation* (3.3)

The problem of satisfying requirement (1) of Theorems 1 and 2 can be solved effectively by a method due to Hotelling [5]. It may be worth noting that this method yields two sets of constants, $a_1, a_2, \cdots, a_l$ and $-a_1, -a_2, \cdots, -a_l$ both maximizing $|\rho(X, U)|$ but leading to values of $\rho(X, U)$ with opposite signs Nevertheless the choice between $a_1, a_2, \cdots, a_l$ and $-a_1, -a_2, \cdots, -a_l$ and the determination of τ are unique for any given m, since (3.1) has a solution for τ only if sign ρ = sign m

5. Linear truncation of $(X, Y_1, Y_2, \cdots, Y_l)$ to the set $\sum_{j=1}^{l} a_j Y_j \geq \tau$ for given expectation of X after truncation, minimizing the variance of X after truncation. It may be of practical interest to choose $a_1, a_2, \cdots, a_l$ and τ so that, with the notations and under the assumptions of section 4, the expectation of X after truncation becomes equal to a given number m, and the variance after truncation is minimum.

THEOREM 3. *To truncate $(X, Y_1, Y_2, \cdots, Y_l)$ linearly in $Y_1, Y_2, \cdots, Y_l$ so that the expectation after truncation has the given value m and that, under this condition, the variance of X after truncation becomes as small as possible, it is necessary and sufficient to satisfy the conditions* (1) *and* (2) *of Theorem* 1.

The proof of this theorem follows from section 3 and the following lemma:

LEMMA 3. *For fixed $E(X) = m$, the variance $\sigma^2(X)$ after truncation is a strictly decreasing function of the absolute value of ρ for $0 < |\rho| \leq 1$*

PROOF: According to (2.3) we have

$$\sigma^2(X) = 1 + m(\rho\tau - m).$$

Differentiating with regard to ρ and using (3.4) we have

$$\frac{d\sigma^2(X)}{d\rho} = m\left(\tau + \rho\frac{d\tau}{d\rho}\right) = m\,\frac{\tau[\lambda(\tau) - \tau] - 1}{\lambda(\tau) - \tau}.$$

For $\tau < 0$ this clearly is <0 For $\tau \geq 0$ inequality (1.7) yields

$$\tau[\lambda(\tau) - \tau] - 1 \leq \tfrac{1}{2}(\tau\sqrt{4 + \tau^2} - 3\tau^2 - 2)$$
$$\leq \tfrac{1}{2}[\tau(2 + \tau) - 3\tau^2 - 2] = \tau(1 - \tau) - 1,$$

and this is < 0 for $\tau \geq 0$. Together with (1.6), this proves that

$$\frac{\tau[\lambda(\tau) - \tau] - 1}{\lambda(\tau) - \tau} < 0$$

for all τ, and hence according to (3.1)

$$\operatorname{sign}\frac{d\sigma^2(X)}{d\rho} = -\operatorname{sign} m = -\operatorname{sign}\rho.$$

It may be conjectured that the sign of $d\sigma^2(X)/d\rho$ is opposite to that of ρ also in the case when $\sigma^2(X)$ is the variance after truncation minimized under condition (3.3). This would lead to a theorem stating that the same choice of $a_1, a_2, \cdots, a_l$ and τ which according to Theorem 2 makes the α-quantile after truncation equal to the given number X_α and minimizes the rejected part of the original population, will also minimize the variance of X after truncation.

REFERENCES

[1] Z. W. BIRNBAUM, E. PAULSON AND F. C. ANDREWS, "On the effect of selection performed on some coordinates of a multi-dimensional population", *Psychometrika*, Vol. 15 (1950)

[2] R. D. GORDON, "Values of Mill's ratio of area to bounding ordinate of the normal probability integral for large values of the argument", *Annals of Math. Stat.*, Vol 12 (1941), pp. 364–366.

[3] Z. W BIRNBAUM, "An inequality for Mill's ratio", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 245–246.

[4] K. PEARSON, *Tables for Statisticians and Biometricians, Part II*, 1st ed., Cambridge Univ. Press, 1931, Tables VIII and IX.

[5] H. HOTELLING, "Relations between two sets of variates", *Biometrika*, Vol. 28 (1936), pp 321–377.

NOTES

This section is devoted to brief research and expository articles and other short items.

EXTENSION OF A THEOREM OF BLACKWELL[1]

BY E. W. BARANKIN

University of California, Berkeley

1. Introduction. In [1] (§1) the author has announced, as bearing on the results there, that Blackwell's method [2] of uniformly improving the variance of an unbiased estimate by taking the conditional expectation with respect to a sufficient statistic, is in fact similarly effective on every absolute central moment of order $s \geqq 1$. Our purpose here is to establish this. In addition, the equality condition (null improvement of the moment) is presented in terms of a primitive property of the estimate. The asserted uniform diminution of the s-th moments for a family W of distributions is, as in the case $s = 2$, a twice removed consequence of the fundamental fact for a single distribution that the absolute s-th power of the conditional expectation of a measurable function is almost everywhere (a.e.) not greater than the conditional expectation of the absolute s-th power of the function This is the substance of the theorem below. The second corollary then states the result for unbiased estimates

2. Preliminaries. Let Ω be a space of points x; $\mathfrak{F}$, a σ-field of subsets of Ω, and μ, a probability measure on $\mathfrak{F}$. Let t be a function on Ω onto a space Γ of points τ; $\mathfrak{T}^\Gamma$ a σ-field of subsets of Γ; and $\mathfrak{T}$—a sub-σ-field of $\mathfrak{F}$—the inverse of $\mathfrak{T}^\Gamma$ under t A set in $\mathfrak{T}^\Gamma$ will be denoted by A^Γ, where A is its inverse under t. Let ν denote the measure on $\mathfrak{T}^\Gamma$ defined by $\nu(A^\Gamma) = \mu(A)$.

If f is a real-valued,[2] $\mathfrak{F}$-measurable, μ-integrable function on Ω, we denote by $E(f \mid \cdot)$ the conditional expectation of f with respect to t Corresponding to any particular function h on Γ (as, for example, $E(f \mid \cdot)$) we define the function h^* on Ω by

$$h^*(x) = h(\tau), \qquad t(x) = \tau.$$

The qualification "essentially" prefixing a statement will mean that with the possible exception of a set of points of measure 0, that statement holds true.

The following two simple lemmas enable us to present the conditions for equality, in the results below, in terms of the elementary characteristics of the function f.

---

[1] This note was prepared under O N. R contract

[2] With no changes in this note, and only minor changes in [1], the results we have set forth concerning unbiased estimation pertain as well to complex-valued functions.

LEMMA 1. *A necessary and sufficient condition that* $\operatorname{sgn} f(x) = \operatorname{sgn} E^*(f \mid x)$ a.e. (μ) *is that* $\operatorname{sgn} f$ *be essentially a function of* t.

The necessity of the condition is clear. To prove sufficiency, let f' be a function on Ω which is a.e. equal to f, and such that $\operatorname{sgn} f'$ is an (unqualified) function of t. Now if $\operatorname{sgn} f'(x) = \operatorname{sgn} E^*(f \mid x)$ does not hold a.e. (μ), then there is a $\mathfrak{T}$-set, A, of positive measure, such that, for example, for $x \in A, f'(x) > 0$ while $E^*(f \mid x) \leqq 0$. We then have the contradiction

$$0 < \int_A f' \, d\mu = \int_A f \, d\mu = \int_A E^*(f \mid \cdot) \, d\mu \leqq 0.$$

LEMMA 2. *A necessary and sufficient condition that* $f(x) = E^*(f \mid x)$ a.e. (μ) *is that* f *be essentially a function of* t.

Again the necessity is obvious. To show sufficiency, let f' be a function on Ω which is a.e. equal to f, and is an (unqualified) function of t. Define h on Γ by

$$h(\tau) = f'(x), \qquad t(x) = \tau.$$

Then $h^* = f'$, and we have

$$\int_A f \, d\mu = \int_A f' \, d\mu = \int_{A\Gamma} h \, d\nu, \qquad A \in \mathfrak{T}.$$

But this implies that $h(\tau) = E(f \mid \tau)$ a.e. (ν), and therefore $f(x) = E^*(f \mid x)$ a.e. (μ), as was to be shown.

3. **Results.** For a proof of the Hölder inequality that we use in establishing the following theorem, we refer the reader to [3] (p. 233).

THEOREM.[3] *Let* $s \geqq 1$. *Then for almost all* $(\mu)x$,

$$|E^*(f \mid x)|^s \leqq E^*(|f|^s \mid x).^4 \tag{1}$$

Equality holds a.e.

(i) *for* $s = 1$, *if and only if* $\operatorname{sgn} f$ *is essentially a function of* t;

(ii) *for* $s > 1$, *if and only if* f *is essentially a function of* t.

PROOF: Consider first the case $s = 1$. Let

$$S = \{x \in \Omega \mid E^*(f \mid x) > 0\},$$
$$S' = \Omega - S.$$

Then, for any $A \in \mathfrak{T}$,

$$\int_A |E^*(f \mid \cdot)| \, d\mu = \int_{SA} E^*(f \mid \cdot) \, d\mu - \int_{S'A} E^*(f \mid \cdot) \, d\mu$$
$$= \int_{SA} f \, d\mu - \int_{S'A} f \, d\mu \leqq \int_A |f| \, d\mu = \int_A E^*(|f| \mid \cdot) \, d\mu.$$

---

[3] The proof we present here was suggested by the referee, and is much shorter than our own.

[4] For $s = 1$ this inequality was used by Doob in "Regularity properties of certain families of chance variables", *Trans. Amer. Math. Soc.*, Vol. 47 (1940), pp 455–486 (Theorem 0.2).

Since A is arbitrary, we have the result (1) with $s = 1$. It is clear that the equality sign holds a.e. (μ) if and only if, except possibly for a set of measure 0, f is positive on S and non-positive on S'; that is, if and only if $\operatorname{sgn} f(x) = \operatorname{sgn} E^*(f \mid x)$ a.e. (μ). Applying Lemma 1, we have the equality condition as stated in the theorem.

Now let $s > 1$. To establish (1) it will suffice, by virtue of what has already been proved for $s = 1$, to consider $f \geqq 0$ a.e. (μ) We may then argue as follows. Unless (1) holds a.e., there is a $\mathfrak{T}$-set, R, of positive measure, and numbers $a > b \geqq 0$ such that for $x \in R$,

$$[E^*(f \mid x)]^s \geqq a,$$

and

$$E^*(f^s \mid x) \leqq b.$$

But then, with an application of the Holder inequality we meet a contradiction. For,

$$\begin{aligned} a[\mu(R)]^s &\leqq \left\{\int_R E^*(f \mid \cdot)\, d\mu\right\}^s = \left\{\int_R f\, d\mu\right\}^s \\ &\leqq \int_R f^s\, d\mu \cdot [\mu(R)]^{s-1} = \int_R E^*(f^s \mid \cdot)\, d\mu \cdot [\mu(R)]^{s-1} \\ &\leqq b[\mu(R)]^s, \end{aligned}$$

which contradicts $a > b$. Thus, (1) is proved in general.

If $f(x) = E^*(f \mid x)$ a.e. (μ), it is readily proved by a direct argument that then equality holds in (1) a.e. (μ). Conversely, suppose equality in (1) holds a.e. Then we have, in fact, a.e.,

$$(2) \qquad | E^*(f \mid x) | = E^*(| f | \mid x),$$

and

$$(3) \qquad [E^*(| f | \mid x)]^s = E^*(| f |^s \mid x).$$

For brevity, denote the function $E^*(| f | \mid \cdot)$ by v. Since f vanishes at almost all points where v vanishes, we may write $| f | = w \cdot v$, where

$$w(x) = \begin{cases} 1, & v(x) = 0, \\ | f(x) | / v(x), & v(x) > 0. \end{cases}$$

(If v vanishes almost everywhere, we are through.) For any $\mathfrak{T}$-measurable, real-valued function, u, on Ω, we have

$$(4) \qquad \int_\Omega u \cdot v\, d\mu = \int_\Omega u \cdot v \cdot w\, d\mu,$$

when either of these integrals exists (cf. [4], p. 50, eq. (15)). Similarly, and taking account of the equality assumption (3) we have

$$\int_\Omega u \cdot v^s \, d\mu = \int_\Omega u \cdot v^s \cdot w^s \, d\mu. \tag{5}$$

In particular, consider the two functions

$$u_1(x) = \begin{cases} 1/v(x), & v(x) > 0, \\ 0, & v(x) = 0, \end{cases}$$

and

$$u_2(x) = \begin{cases} 1/[v(x)]^s, & v(x) > 0, \\ 0, & v(x) = 0. \end{cases}$$

If

$$S_0 = \{x \in \Omega \mid v(x) > 0\},$$

it is seen that u_1 taken in conjunction with (4), and u_2 taken in conjunction with (5), bring out

$$\int_{S_0} w \, d\mu = \int_{S_0} w^s \, d\mu = \mu(S_0).$$

From this it follows (e g , by the equality condition attending the Holder inequality) that $w(x) = 1$ a e. in S_0 Hence $|f(x)| = v(x)$ a.e. in Ω. Therefore, by (2), $|f(x)| = |E^*(f \mid x)|$ a.e. But (2) also implies, as already shown, $\operatorname{sgn} f(x) = \operatorname{sgn} E^*(f \mid x)$ a.e. Thus, finally, we have $f(x) = E^*(f \mid x)$ a.e. Now apply Lemma 2, and the proof of the theorem is complete

COROLLARY 1. *Let $s \geq 1$, and let g_0 denote the expectation of f. Then*

$$\int_\Omega | E^*(f \mid \cdot) - g_0 |^s \, d\mu \leqq \int_\Omega | f - g_0 |^s \, d\mu. \tag{6}$$

Equality holds

(i) *for $s = 1$, if and only if* sgn $[f - g_0]$ *is essentially a function of t,*

(ii) *for $s > 1$, if and only if f is essentially a function of t.*

This result expresses the domination over the s-th absolute central moment of the conditional expectation of f by the corresponding moment of f itself. It follows almost immediately from the theorem when we write (6) in the form

$$\int_\Omega | E^*(f - g_0 \mid \cdot) |^s \, d\mu \leqq \int_\Omega E^*(| f - g_0 |^s \mid \cdot) \, d\mu. \tag{7}$$

Thus, from the theorem we know that the integrand of the left-hand side of (7) is a e. $\leqq$ the integrand on the right. Hence (7) holds. Equality in (7) holds then if and only if the integrands are a e. equal. The theorem therefore directly provides the equality conditions as stated.

Let $W = \{\mu_\theta, \theta \in \Theta\}$ be a family of probability measures on $\mathfrak{F}$; and t, a sufficient

statistic for W (cf. [5], p. 232, §5). Let f be an unbiased estimate of the function g^2 on Θ For each $\mu_\theta \in W$, the conditional expectation, $E_\theta(f \mid \ )$, of f with respect to t is defined. Since conditional expectations are fully determined by conditional probabilities (although, in general, not as usual integrals. Cf. [4], pp 48, 49; also [5], p 230) it follows from the sufficiency of t that there exists a function $E(f \mid \cdot)$, on Γ, with $E_\theta(f \mid \tau) = E(f \mid \tau)$ a.e. (ν_θ) for each $\theta \in \Theta \cdot E^*(f \mid \cdot)$ is again an unbiased estimate of g, and we have

COROLLARY 2. *Let t be a sufficient statistic for the family $W = \{\mu_\theta, \theta \in \Theta\}$; and f, an unbiased estimate of g. For $s \geqq 1$, and each $\theta \in \Theta$,*

$$\int_\Omega | E^*(f \mid \cdot) - g(\theta) |^s d\mu_\theta \leqq \int_\Omega | f - g(\theta) |^s d\mu_\theta .$$

Equality holds

(i) *for $s = 1$, if and only if* sgn $[f - g(\theta)]$ *is essentially (μ_θ) a function of t;*

(ii) *for $s > 1$, if and only if f is essentially (μ_θ) a function of t.*

REFERENCES

[1] E W. BARANKIN, "Locally best unbiased estimates", *Annals of Math Stat.*, Vol. 20 (1949), pp 477–501.

[2] D BLACKWELL, "Conditional expectation and unbiased sequential estimation", *Annals of Math. Stat.*, Vol. 18 (1947), pp. 105–110

[3] L. M. GRAVES, *The Theory of Functions of Real Variables*, McGraw-Hill, 1946

[4] A KOLMOGOROFF, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Vol. 2 (1933).

[5] P. R HALMOS AND L. J. SAVAGE, "Application of the Radon-Nikodym Theorem to the theory of sufficient statistics", *Annals of Math. Stat* , Vol. 20 (1949), pp. 225–241.

NOTE ON CONSISTENT ESTIMATES OF THE LINEAR STRUCTURAL RELATION BETWEEN TWO VARIABLES[1]

BY ELIZABETH L SCOTT

University of California, Berkeley

1. Introduction. The purpose of this note is to present another case in which the structural linear relation between two observable random variables may be consistently estimated. Of the recent papers on this subject I wish to mention the paper by Wald [1], which contains a history of the work done on the problem, and the more recent paper by Housner and Brennan [2]. Also relevant is the important result due to Reiersøl [3], [4]

2. Statement of problem. Assume that the two observable random variables x and y have the structure

[1] Paper prepared with partial support of the Office of Naval Research

The results summarized were presented in a discussion held at the Cleveland Meeting of the Institute of Mathematical Statistics, December, 1948

$$\begin{cases} x = \xi + u \\ y = \alpha + \beta\xi + v, \end{cases} \tag{1}$$

where α and β are unknown parameters to be estimated, and ξ, u and v are completely independent random variables. The latter two variables, interpreted as the random errors of measurement, are assumed to vary normally about zero with unknown variances σ_1^2 and σ_2^2, respectively.

An increasing number n of completely independent pairs of simultaneous values of x and y are to be observed

$$(x_i, y_i), \qquad i = 1, 2, \cdots, n, \tag{2}$$

so that each pair (x_i, y_i) corresponds to a value ξ_i of the unobservable random variable ξ which is independent of the value ξ_j of ξ corresponding to any other pair (x_j, y_j), $i \neq j$.

It is well known that if the distribution of ξ is normal then the parameters α, β, σ_1 and σ_2 are unidentifiable. Reiersøl proved [4] that these parameters are identifiable in all other cases. Wald and Housner and Brennan found consistent estimates of these parameters assuming that, although the particular values of ξ are not known exactly, a certain amount of knowledge concerning the values of ξ is available. The present note gives a method for obtaining a consistent estimate of β, which is the key to the problem of estimating the four parameters, for the case where it is known that a specified central moment of the distribution of ξ exists and differs from that of the normal distribution.

Since work on this subject continues, the present brief note deals particularly with the simplest case, when one of the odd central moments of ξ exists and differs from the "normal" value, zero. It will be observed that the hypotheses made here are of entirely different character from those adopted by other writers. The present note postulates knowledge concerning a moment of the distribution of ξ, whereas the papers quoted postulate some knowledge of the particular values assumed by ξ. The method adopted was suggested by a remark made by Neyman [5] in 1936.

3. Preliminary theorems. Let

$$x_{\cdot} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad y_{\cdot} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3}$$

and let b be an arbitrary real number.

THEOREM 1: *If μ_3, the third central moment of ξ, exists then the arithmetic mean*

$$F_{n,1}(b) = \frac{1}{n}\sum_{i=1}^{n} [y_i - y_{\cdot} - b(x_i - x_{\cdot})]^3 \tag{4}$$

converges in probability to

$$(\beta - b)^3\mu_3. \tag{5}$$

PROOF. Simple algebra gives

$$
\begin{aligned}
F_{n,1}(b) = {} & (\beta - b)^3 \frac{1}{n} \sum_{i=1}^{n} (\xi_i - \xi_.)^3 \\
& + 3(\beta - b)^2 \frac{1}{n} \sum_{i=1}^{n} (\xi_i - \xi_.)^2 [v_i - v_. - b(u_i - u_.)] \\
& + 3(\beta - b) \frac{1}{n} \sum_{i=1}^{n} (\xi_i - \xi_.)[v_i - v_. - b(u_i - u_.)]^2 \\
& + \frac{1}{n} \sum_{i=1}^{n} [v_i - v_. - b(u_i - u_.)]^3.
\end{aligned} \tag{6}
$$

It is obvious that further expansion will express $F_{n,1}(b)$ in terms of averages of the type

$$\frac{1}{n} \sum_{i=1}^{n} \xi_i^p u_i^q v_i^r, \tag{7}$$

with $p + q + r \leqq 3$. Since all the terms over which each average is taken are completely independent, follow the same law and possess finite expectations, the familiar theorem of Khintchine assures that, as n is increased, each average (7) tends in probability to its expectation. Using Slutsky's theorem (see Cramér [6], p. 255), we conclude that $F_{n,1}(b)$ tends in probability to the limit obtained by replacing each average in the expansion (6) by its expectation and then letting $n \to \infty$. The computations are easy and give

$$\lim_{n \to \infty} pF_{n,1}(b) = (\beta - b)^3 \mu_3 . \tag{8}$$

Q.E.D.

Let $\{X_n\}$ denote a sequence of observable random variables (multivariate or not) such that the distribution function of X_n depends on the parameters θ_i with $a_i < \theta_i < b_i$, $i = 1, 2, \cdots, s$. Furthermore, let λ denote a real variable and $\{\phi_n(X_n, \lambda)\}$ a sequence of functions of the arguments X_n and λ defined for all possible values of X_n and for all values of λ within the limits $a_1 \leqq \lambda \leqq b_1$.

THEOREM 2: *If the sequence of functions $\{\phi_n(X_n, \lambda)\}$ has the following properties:*

(i) *whatever be the true values θ_1', θ_2', $\cdots$, θ_s' of the parameters θ_i within the limits $a_i < \theta_i' < b_i$, $i = 1, 2, \cdots, s$, as n is increased, the sequence $\{\phi_n(X_n, \lambda)\}$ tends in probability to a function $f(\lambda, \theta_1')$ of arguments λ and θ_1' only.*

(ii) *whatever be $\delta > 0$, there exist in (a_1, b_1) two numbers λ_1 and λ_2, each differing from θ_1' by less than δ and such that the product $f(\lambda_1, \theta_1') f(\lambda_2, \theta_1')$ is negative,*

(iii) *for every n and every possible value x_n of X_n, the function $\phi_n(x_n, \lambda)$ is continuous with respect to λ for $a_1 \leqq \lambda \leqq b_1$,*

then whatever be $\epsilon > 0$ and $\eta > 0$ there exists a number $N_{\epsilon,\eta}$ such that for $n > N_{\epsilon,\eta}$ the probability that the equation $\phi_n(X_n, \lambda) = 0$ has a root between $\theta_1' - \epsilon$ and $\theta_1' + \epsilon$ exceeds $1 - \eta$

PROOF: Let $\epsilon > 0$ and $\eta > 0$ be two arbitrarily small numbers. Let λ_1 and λ_2 be two numbers such that $\lambda_i \in (a_1, b_1)$ and $| \theta_1' - \lambda_i | < \epsilon$, $i = 1, 2$, and such

that $f(\lambda_1, \theta_1') < 0 < f(\lambda_2, \theta_1')$ Select $N_{\epsilon,\eta}$ so large that for $n > N_{\epsilon,\eta}$ the probability of having simultaneously

(9) $$|\phi_n(X_n, \lambda_i) - f(\lambda_i, \theta_1')| < \tfrac{1}{2}|f(\lambda_i, \theta_1')| \quad \text{for } i = 1, 2$$

differs from unity by less than η. It is clear that if the inequalities (9) are satisfied for any particular value x_n of X_n, then

(10) $$\phi_n(x_n, \lambda_1) < 0 < \phi_n(x_n, \lambda_2)$$

and the continuity of $\phi_n(x_n, \lambda)$ for $\lambda \in (a_1, b_1)$ implies that there exists a number $\lambda(x_n)$ between λ_1 and λ_2 such that $\phi_n(x_n, \lambda(x_n)) = 0$ Obviously $|\theta_1' - \lambda(x_n)| < \epsilon$. Thus, whatever be $\epsilon, \eta > 0$, there exists a number $N_{\epsilon,\eta}$ such that the probability that $\phi_n(X_n, \lambda)$ has a root in the interval $(\theta_1' - \epsilon, \theta_1' + \epsilon)$ exceeds $1 - \eta$ provided $n > N_{\epsilon,\eta}$. This proves Theorem 2.

Theorem 2 is treated as a convenient lemma on which to base the proof of the existence of a consistent estimate of the parameter β in (1). It is obvious, however, that this Theorem has an independent interest of its own.

4. Consistent estimates of the structural parameter β. Referring to the general set-up of the problem of estimating the structural parameter β in (1) and using the notation (2) and (3), we prove the following theorems.

THEOREM 3: *If the third central moment μ_3 of ξ exists and differs from zero, then the equation*

(11) $$F_{n,1}(b) = \frac{1}{n}\sum_{i=1}^{n}[y_i - y_. - b(x_i - x_.)]^3 = 0$$

has a root $\hat{b}$ which is a consistent estimate of β.

PROOF: According to Theorem 1, whatever be b and β, the stochastic limit of $F_{n,1}(b)$ is $(\beta - b)^3\mu_3$ and changes its sign as b passes through the value β. Theorem 2 implies then that whatever be $\epsilon, \eta > 0$, there exists a number $N_{\epsilon,\eta}$ such that for $n > N_{\epsilon,\eta}$ the probability that at least one of the roots of (11) will lie within $\beta - \epsilon$ and $\beta + \epsilon$ is greater than $1 - \eta$. This proves the theorem.

Generally, let μ_m denote the m^{th} central moment of ξ.

THEOREM 4: *If the distribution of ξ has moments up to and including order $2m + 1$ and if at least one of the first m odd central moments μ_{2k+1} differs from zero, $k = 1, 2, \cdots, m$, then the equation*

(12) $$F_{n,m}(b) = \frac{1}{n}\sum_{i=1}^{n}[y_i - y_. - b(x_i - x_.)]^{2m+1} = 0$$

has a root $\hat{b}$ which is a consistent estimate of β.

PROOF: The proof of Theorem 4 exactly follows the lines of that of Theorem 3. Using (1), (2) and (3), we write

(13) $$F_{n,m}(b) = \sum_{k=0}^{2m+1} C_{2m+1}^{k}(\beta - b)^k \left\{\frac{1}{n}\sum_{i=1}^{n}(\xi_i - \xi_.)^k[v_i - v_. - b(u_i - u_.)]^{2m+1-k}\right\}.$$

It is easily seen that, as $n \to \infty$, $F_{n,m}(b)$ tends in probability to the limit

$$F_{n,m}(b) \underset{n \to \infty}{\xrightarrow[p]{}} (\beta - b)^3 \psi(\beta - b), \tag{14}$$

where $\psi(\beta - b)$ is a linear combination of even powers of $(\beta - b)$ with at least one coefficient different from zero It follows that the stochastic limit of $F_{n,m}(b)$ changes its sign as b passes through β and the proof is completed by reference to Theorem 2.

Note that the stochastic limit of the first derivative of $F_{n,m}(b)$, evaluated at $b = \beta$, is zero, which is unfortunate. Furthermore, the order of contact of $F_{n,m}(b)$ at $b = \beta$ increases with the order of the first odd central moment of ξ which differs from zero. Therefore, the precision of estimating β may be expected to be better when the low odd central moments are not zero Without narrowing the generality of the case considered, it is difficult to make an evaulation of the precision of the estimates obtained. Thus, for example, the familiar method of evaluating the asymptotic variance requires the knowledge of higher moments of ξ than those considered here For similar reasons, it is thus far impossible to speak of the relative efficiency of the estimates found For this purpose it would be necessary to determine first the measure of the precision of the best estimate whose consistency persists independently of the distribution of ξ provided only that at least one odd central moment differs from zero

Once the consistent estimate $\hat{b}$ of β is obtained, there is no particular difficulty in obtaining consistent estimates of the other parameters

J Neyman has pointed out [7] that Theorem 2 may be used as the basis for a very elementary proof of the consistency of maximum likelihood estimates

REFERENCES

[1] Abraham Wald, "The fitting of straight lines if both variables are subject to error," *Annals of Math. Stat*, Vol. 11, (1940), p 284

[2] G W. Housner and J F Brennan, "The estimation of linear trends," *Annals of Math Stat.*, Vol 19 (1948), p 380.

[3] Olav Reiersøl, "Confluence analysis by means of lag moments and other methods of confluence analysis," *Econometrica*, Vol. 9 (1941), p 1

[4] Olav Reiersøl, "Identifiability of a linear relation between variables which are subject to error," *Cowles Commission Discussion Papers, Statistics*, No 337 (1949)

[5] J. Neyman, *Jour Roy Stat Soc*, Vol 100 (1937), p 56

[6] H Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[7] J Neyman, *First Course in Probability and Statistics*, Vol 2, forthcoming

ON MULTINOMIAL DISTRIBUTIONS WITH LIMITED FREEDOM: A STOCHASTIC GENESIS OF PARETO'S AND PEARSON'S CURVES

By Maria Castellani

University of Kansas City

1. A multinomial law with limited freedom: Distribution functions of statistical equilibrium. We intend to consider here a convenient model of statistical mechanics, which by generalization of an approach used by Cantelli [1], shall give us either Pareto's or Pearson's curves. Let us imagine that N elements ($N > k - 3$) have to be randomly distributed in a set of k continuous intervals t_i ($i = 1, 2, \cdots, k$) in R_1, the "a priori" probability associated with t_i, being p_i, for $\sum_{i=1}^{k} p_i = 1$. Assuming that the elements have no preferences, they move freely under the law of chance taking different configurations $(n_1, n_2, \cdots, n_k)$, with probabilities $P(n_1, n_2, \cdots, n_k)$, n_i being the number of elements placed in t_i and $\sum_{i=1}^{k} n_i = N$. The random variable $Y(t)$ representing the total number of configurations $(n_1, n_2, \cdots, n_k)$, therefore obeys a multinomial law with $k - 1$ degrees of freedom, viz:

$$(1) \qquad P_{[n_1,n_2,\cdots,n_k]} = N! \prod \frac{1}{n!} p_i^{n_i}, \qquad \sum_{i=1}^{k} n_i = N, \qquad \sum_{i=1}^{k} p_i = 1.$$

We shall proceed to admit that the elements are not free, but that they have preference in the choice of a suitable interval. This fact we associate with the assumption that some forces of attraction are made to play in each interval. For the sake of simplicity we shall consider that there are two independent forces, say $\mu(t)$ and $\nu(t)$, whose convenient potential functions are respectively $f(t)$ and $\varphi(t)$, where

$$(2) \qquad \frac{df(t)}{dt} = -\mu(t); \qquad \frac{d(\varphi t)}{dt} = -\nu(t)$$

These potential functions we may, for instance, associate with the significance of a certain quanta whose total is to be distributed among the elements and whose significance must be established by consideration of the particular statistical experiment. It is then admissible, at least in our first approach, to assume these potential quantities to have a total constant magnitude, *viz*, $\sum n_i f(t_i) = H_1$, $\Sigma\, n_i \varphi(t_i) = H_2$, where H_1 and H_2 are appropriate constants. This condition is analogous to the assumption in statistical mechanics of the preservation of energy. This analogy enables us to follow classical methods. We shall call our method the method of "intervals of energy." Let us say that our system reaches its canonic state when $P_{(n_1,n_2,\cdots,n_k)}$ is a maximum [2]. When this state occurs with a probability close to the value one, we may say it is in statistical equilibrium. It is well known that $P_{(n_1,n_2,\cdots,n_k)}$ reaches its maximum when

$$(3) \qquad \delta P_{(n_1,n_2,\cdots,n_k)} = 0 \quad \text{or} \quad \delta \log P_{(n_1,n_2,\cdots,n_k)} = 0$$

Performing as usual, for example as in [3], we ultimately obtain:

$$n_i = Np_i e^{-a-bf(x_i)-c\varphi(x_i)}, \tag{4}$$

where a, b, and c are arbitrary constants.

If N is sufficiently large, n_i/N may be considered a probability, and precisely the probability for $Y(t)$ to score n_i times the value t_i when the canonical state is reached. The problem may then be extended assuming that a continuous function may interpolate the discrete values $n_1, n_2, \cdots, n_k$. Putting $y = n_i/N$, assuming for the sake of simplicity that $f(t)$ and $\varphi(t)$ along with their derivatives are continuous functions of t, and grouping these constants into a single K, formula (4) becomes:

$$y = Ke^{-bf(t)-c\varphi(t)}, \tag{5}$$

then:

$$\frac{d \log y}{dt} = \frac{1}{y}\frac{dy}{dt} = -b\frac{df(t)}{dt} - c\frac{\varphi(t)}{dt} = -b\mu(t) - c\nu(t). \tag{6}$$

Equation (6) is a generalization of the familiar Pearson differential equation which generates his system of curves. It is obvious that (6) may determine a large set of frequency curves, depending on the form of $f(t)$ and $\varphi(t)$.

The above analysis may be extended to any number of acting forces provided they are less than $k - 1$ in number

2. Stochastic genesis of the Pareto and Pearson curves. We shall next show how the Pareto and Pearson curves belong to this family of frequency curves

The Pearsonian system of curves is derived by comparing its differential equation with (6) to determine in these the most suitable functions for $\mu(t)$ and $\nu(t)$ Thus,

$$\frac{1}{y}\frac{dy}{dt} = \frac{t+\alpha}{\beta_1 + \beta_2 t + \beta_3 t^2} = -b\mu(t) - c\nu(t) \tag{7}$$

Corresponding to the decomposition into partial fractions of the middle term, we have two sets of curves. When

$$(\beta_1 + \beta_2 t + \beta_3 t^2) = \beta_3(t-\gamma_1)(t-\gamma_2)$$

and γ_1, γ_2 are real numbers, then

$$\begin{aligned} -b\mu(t) &= \frac{\gamma_1+\alpha}{\beta_3(\gamma_1-\gamma_2)(t-\gamma_1)} = \frac{p}{t-\gamma_1}, \\ -c\nu(t) &= \frac{\gamma_2+\alpha}{\beta_3(\gamma_2-\gamma_1)(t-\gamma_2)} = \frac{q}{t-\gamma_2}, \end{aligned} \tag{8}$$

p and q being suitable grouping constants

Under these assumptions two forces are acting in each "class of energy", each one being proportional to the distance of the interval from some origin.

Substituting (8) into (6) and integrating, we obtain corresponding to the first of (5), after grouping the exponential constants into K

$$y = K(t - \gamma_1)^p(t - \gamma_2)^q,$$

where K, γ_1, γ_2, p, q also have the significance of statistical constants according to which we obtain Pearson's curves of Type I or VI.

When $\beta_3 = 0$, we have by the same process:

$$-b\mu(t) = \frac{1}{\beta_2} = q_2, \qquad -c\nu(t) = \frac{\alpha - \beta_1/\beta_2}{\beta_1 + \beta_2 t} = \frac{q_1}{p + t}$$

Hence, by grouping together the statistical constants under K, p, q_1, q_2, we obtain:

$$y = K(p + t)^{q_1}e^{q_2 t},$$

which is a Pearson curve of Type III.[1] In each class interval two forces are acting; one is constant and the other is inversely proportional to the distance of the interval from a fixed origin

We obtain a Pareto curve when in (8) either p or q is zero. Under the indicated assumptions the Pareto income distribution curve appears in a new light. If the acting forces are reduced to one, and this one force is inversely proportional to the distance of the interval from some origin, the Pareto curve represents a special case of the Pearsonian curve

In (7), we now consider the decomposition of the Pearson function for the case where the denominator does not have real roots This decomposition may be indicated as follows:

$$\frac{t + \alpha}{\beta_1 + \beta_2 t + \beta_3 t^2} = \frac{t + \beta_2/2\beta_3}{\beta_3\{(t + \beta_2/2\beta_3)^2 + \beta_1/\beta_3 - \beta_2^2/4\beta_3^2\}} + \frac{\alpha - \beta_2/2\beta_3}{\beta_3\{(t + \beta_2/2\beta_3)^2 + \beta_1/\beta_3 - \beta_2^2/4\beta_3^2\}}.$$

Setting

$$\frac{\beta_2}{2\beta_3} = p_1, \qquad \alpha - \frac{\beta_2}{2\beta_3} = p_2, \qquad \frac{\beta_1}{\beta_3} - \frac{\beta_2^2}{4\beta_3} = q,$$

$$-b\mu(t) = \frac{t + p_1}{\beta_3\{(t + p_1)^2 + q\}}, \qquad -c\nu(t) = \frac{p_2}{\beta_3\{(t + p_1)^2 + q}.$$

[1] A L Bowley has found in his well-known analysis of food expenditures of urban families, that the distribution of weekly family expenditures can be best expressed by a Pearson curve of Type III This is not surprising, since it is exactly a case where we can assume the joint effect of a constant factor and another factor acting in inverse proportion to the interval (again in the sense of the distance from a suitable origin) The constant factor in our case is the human need of food, while the factor acting inversely to the interval can be taken as a response to prices See [4]

These we may interpret as forces of the Newtonian type By grouping the statistical constants appropriately under K, p_1, q, m_1, m_2, m_3, we derive from (7) the following equation.

$$y = k\{(t + p_1)^2 + q\}^{m_1} e^{m_2 \tan^{-1}((t+p_1)/m_3)}.$$

This is the familiar Pearson curve of Type IV.

Other distributions of the same family can be easily found by the same method.

3. The frequency curves and their statistical equilibrium. The conclusive step in this analysis is in finding the probability of the most likely configuration. By generalizing a process of statistical mechanics first used by Castelnuovo [5], we assume any configuration $(n_1', n_2', \cdots, n_k')$ slightly different from the most probable $(n_1, n_2, \cdots, n_k)$ (the canonical configuration). Setting

$$n_i' = \alpha_i + n_i \qquad (i = 1, 2, \cdots, k),$$

we have by conditions (1) and (3):

$$\sum_{i=1}^{k} \alpha_i = 0, \qquad \sum_{i=1}^{k} \alpha_i f(t_i) = 0, \qquad \sum_{i=1}^{k} \alpha_i \varphi(t_i) = 0, \tag{9}$$

$$P_{[n_1', n_2', \cdots, n_k']} = N! \prod \frac{1}{n_i'!} p_i^{n_i'}.$$

The sum of the values of $P_{(n_1', n_2', \ldots, n_k')}$ will give us the total probability of scoring a n_i' slightly different from n_i. Let us designate by Π the total probability of having $P_{(n_1', n_2', \cdots, n_k')}$ satisfying all above conditions. By following Castelnuovo's method [2], [5], we obtain:

$$\prod = \sum P_{[n_1', n_2', \cdots, n_k']} \simeq P_{[n_1, n_2, \cdots, n_k]} \sum \exp\left[-\tfrac{1}{2} \sum_{i=1}^{k} \frac{\alpha_i^2}{n_i}\right].$$

We determine all integral sets of n_i compatible with (9); and with a condition of size

$$\sum_{i=1}^{k} \frac{\alpha_i^2}{n_i} \leq 2\mu_0.$$

By a well-known process [2], [5] for any u_0

$$\prod = \frac{1}{\Gamma\left(\dfrac{k-3}{2}\right)} \int_0^{u_0} u^{(k-5)/2} e^{-u}\, du.$$

This is the familiar Chi-square distribution function with $(k - 3)$ degrees of freedom By considering u_0 as increasing with N, we can conclude that

$$\lim_{N \to \infty} \prod = 1.$$

The state of maximum likelihood has a real significance only if it is almost certain that we will obtain either such a state or any one practically equivalent to it.

This occurs when the state of maximum probability has little chance to change; it is a so-called *stationary state* or state of *statistical equilibrium* It would mean a great deal if we could be able to say through how many states the statistical phenomena must pass before attaining its equilibrium, or in other words, whether the ergodic hypothesis of the kinetic theory of gas applies to certain social or economic phenomena. We will not go further into this now, the results obtained here must be considered as an initial exploratory step, which does permit us, however, to end with the following conclusive statement:

If N elements, provided N is large enough, are distributed at random in k class "intervals of energy", it is highly probable that they will approach a configuration of statistical equilibrium, a distribution of maximum probability. Pareto's and Pearson's curves represent special configurations of statistical equilibrium in a stochastic system.

REFERENCES

[1] F P. CANTELLI, "Sulle deduzioni delle leggi di frequenza da considerazioni di probabilità," *Metron*, Vol 1 (1921), N. 3

[2] G. CASTELNUOVO, *Calcolo della Probabilità*, Roma, 1919.

[3] R B LINDSAY, *Introduction to Physical Statistics*, New York, 1941

[4] A L. BOWLEY, *Elements of Statistics*, London, 1926

[5] J. L COOLIDGE, *An Introduction to Mathematical Probability*, Oxford, London, 1925.

ON THE COMPLETELY UNBIASSED CHARACTER OF TESTS OF INDEPENDENCE IN MULTIVARIATE NORMAL SYSTEMS

BY R. D. NARAIN

Indian Council of Agricultural Research

1. Introductory. To prove the unbiassed character of likelihood ratio tests like the test of significance of the multiple correlation coefficient or Hotelling's T^2 test, Daly [1] used the non-null frequency distributions of these test criteria. This leads to obvious difficulties when tackling the general regression problem and the test of independence of several sets of variates, and Daly [1] has shown only their locally unbiassed character.

This paper demonstrates an approach which does not require an explicit knowledge of the frequency distribution of the test criteria and it has been possible to prove that the likelihood ratio test for the general regression problem and the Wilks' criterion for independence of sets of variates are completely unbiassed. The argument proceeds in a chain, the unbiassedness of the Wilks' criterion following ultimately from the unbiassedness of the t-test. The link up has been achieved by working with a chain of conditional distribution densities, a principle employed earlier by the author [3], [4] in presenting a unified distribution theory of the common statistical coefficients relevant to normal theory.

2. The t-test. As the simplest demonstration of the procedure which is applicable generally, consider the t-test for the significance of the mean of a normal population. Let the frequency function of a sample of size n be

$$(1) \qquad (2\pi V)^{-(n/2)} \exp\left[-\frac{1}{2V}\sum_{i=1}^{n}(x_i - m)^2\right].$$

The region $W - w$ complementary to the critical region w for testing the hypothesis

$$m = 0$$

is given by

$$\bar{x}^2 \leq k^2\chi^2,$$

where k is a positive constant depending on the size of w and

$$n\bar{x} = \sum_{i=1}^{n} x_i,$$

$$\chi^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2.$$

We write

$$(2) \qquad I(m) = \int_0^\infty \left[\int_{-k\chi}^{k\chi} e^{-((n/2V)(\bar{x}-m))^2}\, d\bar{x}\right] f(\chi^2)\, d(\chi^2),$$

where

$$f(\chi^2)\, d(\chi^2)$$

is the frequency function of χ^2 which is distributed independently of $\bar{x}$. To show that the test is completely unbiassed is equivalent to showing that

$$I(m) \leq I(0) \text{ for all values of } m.$$

We have

$$\frac{\partial I}{\partial m} = \int_0^\infty \{e^{-(n/2V)(K\chi+m)^2} - e^{-(n/2V)(K\chi-m)^2}\}\, f(\chi^2)\, d(\chi^2)$$

which is positive or negative according as m is negative or positive. Therefore

$$I(m) \leq I(0).$$

3. The E^2 and R^2 tests. Let the frequency function of n observations of a random variate x_p be

$$(3) \qquad (2\pi V)^{-(n/2)} \exp\left[-\frac{1}{2V}\sum_{i=1}^{n}\left(x_{ip} - \sum_{r=1}^{p-1}\beta_r x_{ir}\right)^2\right]\prod_i dx_{ip}.$$

With the usual notation for partial variates in regression analysis, the critical region w based on the likelihood ratio test for the hypothesis

$$0 = \beta_m = \beta_{m+1} = \cdots = \beta_{p-1}, \qquad m \leq p - 1,$$

is given by

$$1 - E^2 = \frac{\sum_i x^2_{ip\,(12\cdots p-1)}}{\sum_i x^2_{ip\,(12\cdots m-1)}} \leq \text{a positive constant}$$

It can be shown [2], [3] that this ratio can be expressed in the form

$$1 - E^2 = \frac{\chi^2}{\chi^2 + \sum_{r=m}^{p-1} z_r^2},$$

where the frequency function of χ^2 and the z_r is

$$(4) \qquad \frac{(2\pi V)^{-(n/2)}\,(\pi)^{-(p-1)/2}}{\Gamma\left(\frac{n-p+1}{2}\right)} \cdot \exp\left[-\frac{1}{2V}\left\{\chi^2 + \sum_{r=1}^{p-1} (z_r - \eta_r)^2\right\}\right] (\chi^2)^{((n-p-1)/2)}\, d(\chi^2) \prod_r dz_r$$

The hypothesis to be tested then becomes

$$0 = \eta_m = \eta_{m+1} = \cdots = \eta_{p-1}.$$

The region $W - w$ complementary to w is given

$$\sum_{r=m}^{p-1} z_r^2 \leq k\chi^2,$$

where k is a positive constant determined by the size of w. Denote by $I(\eta_{p-1}, \eta_{p-2}, \cdots \eta_m)$ the integral of (4) over the region $W - w$. Differentiating I with respect to η_{p-1}, performing the integration with respect to z_{p-1} and arguing exactly as in section 2 above we obtain

$$I(\eta_{p-1}, \eta_{p-2}, \cdots, \eta_m) \leq (0, \eta_{p-2}, \eta_{p-3} \cdots \eta_m).$$

Note that z_{p-2} is distributed independently of z_{p-1}. Therefore starting with $\eta_{p-1} = 0$ in (4) and considering the integration with respect to z_{p-2} first, we obtain as before $I(0, \eta_{p-2} \cdots \eta_m) \leq I(0, 0, \eta_{p-3} \cdots \eta_m)$ and thus finally $I(\eta_{p-1}, \eta_{p-2}, \cdots \eta_m) \leq I(0, 0, \cdots 0)$, which proves the completely unbiassed character of the E^2-test. The test of significance of the multiple correlation coefficient with any number of the predicting variates being fixed or random may be considered as a corollary to the above. We have only to multiply the frequency function (3) by a factor dF representing the frequency function of the random predicting variates (which need not be necessarily normal). This does not affect either the test criterion or the arguments showing its unbiassedness. The test of significance of the multiple correlation coefficient is thus completely unbiassed.

4. The general regression problem. Given the distribution,

$$(5)\qquad (2\pi)^{-\frac{1}{2}pn} \,|\alpha^{rs}|^{n/2} \exp\Big\{-\tfrac{1}{2}\sum_{r,s}\alpha^{rs}\Big[\sum_i \big(x_{ir} - \sum_h \beta_{rh}\, x_{ih}\big) \cdot \big(x_{is} - \sum_j \beta_{sj}\, x_{ij}\big)\Big]\Big\} \times \prod_{i,r} dx_{ir},$$

$$i = 1, 2, \cdots n,$$
$$h = 1, 2, \cdots l, l+1, l+2, \cdots m,$$
$$r, s = m+1, m+2, \cdots p,$$
$$n > p > m \geq l,$$

where the matrix $\| x_{ih} \|$ is of rank m. The hypothesis H to be tested is

$$\beta_{rv} = 0, \qquad \begin{array}{l} r = m+1, m+2, \cdots p, \\ v = l+1, l+2, \cdots m. \end{array}$$

The likelihood ratio test gives the critical region defined by

$$\lambda \equiv \frac{|a_{rs}|}{|a'_{rs}|} \leq \text{ a positive constant},$$

where, with the usual regression notation for partial variates,

$$\begin{aligned} a_{rs} &= \sum_{i=1}^{n} x_{ir\cdot(12\cdots m)}\, x_{is\cdot(12\cdots m)}, \\ a'_{rs} &= \sum_{i=1}^{n} x_{ir\cdot(12\cdots l)}\, x_{is\cdot(12\cdots l)}, \end{aligned} \qquad r, s = m+1, m+2, \cdots p,$$

Now we note that

$$(6)\qquad \lambda = \prod_{r=m+1}^{p} (1 - E_r^2) = (1 - E_p^2) \prod_{r=m+1}^{p-1} (1 - E_r^2),$$

where

$$1 - E_r^2 = \frac{\sum_{i=1}^{n} x^2_{ir\cdot(12\cdots l, l+1, l+2, \cdots m, m+1, \cdots r-1)}}{\sum_{i=1}^{n} x^2_{ir\cdot(12\cdots l, m+1, m+2, \cdots r-1)}}.$$

Since the statistic λ is invariant to linear transformations of the random variates $x_{m+1}, x_{m+2}, \cdots, x_p$ the distribution (5) may be simplified to

$$(7)\qquad \prod_{r=m+1}^{p}\Big[(2\pi V_r)^{-(n/2)} \exp\Big[-\frac{1}{2V_r}\sum_i \big(x_{ir} - \sum_h \beta_{rh}\, x_{ih}\big)^2\Big] \prod_i dx_{ir}\Big],$$

$$i = 1, 2, \cdots n,$$
$$h = 1, 2, \cdots m.$$

Denote by $I(\beta_{pv}, \beta_{p-1,v}, \cdots \beta_{m+1,v})$ the integral of (7) over the region $W - w$

complementary to the critical region w, where β_{rv} in I stands for the entire set of parameters $\beta_{r,t+1}, \beta_{r,t+2}, \cdots \beta_{rm}$. We may first integrate over a subregion of $W - w$ over which $\prod_{r=t+1}^{p-1} (1 - E_r^2)$ has a given value. Using identity (6) and the result of section 3 it follows immediately that

$$I(\beta_{pv}, \beta_{p-1,v}, \cdots \beta_{m+1,v}) \leq I(0, \beta_{p-1,v}, \beta_{p-2,v}, \cdots \beta_{m+1,v}).$$

If $\beta_{pv} = 0$, the distribution of E_p^2 is independent of that of E_{p-1}^2. Hence, starting with $\beta_{pv} = 0$ in (7) and considering the integration for E_{p-1}^2 first, we obtain

$$I(0, \beta_{p-1,v}, \beta_{p-2,v}, \cdots \beta_{m+1,v}) \leq I(0, 0, \beta_{p-2,v}, \cdots \beta_{m+1,v}).$$

Thus finally

$$I(\beta_{pv}, \beta_{p-1,v}, \cdots \beta_{m+1,v}) \leq I(0, 0, \cdots 0),$$

which proves the completely unbiassed character of the test.

5. Test of independence of sets of variates. Consider n observations of q sets of random variates distributed in the multivariate normal form

$$\text{Const} \times \exp \left[-\tfrac{1}{2} \sum \alpha^{rs} \left\{\sum_i (x_{ir} - m_r)(x_{is} - m_s)\right\}\right] \prod_{i,r} dx_{ir},$$

(8) $$\begin{aligned} i &= 1, 2, \cdots n, \\ r &= 1, 2, \cdots l_1, l_1 + 1, l_1 + 2, \cdots l_2, l_2 + 1, \cdots l_3, \cdots l_q, \\ n &> l_q. \end{aligned}$$

Denote by D_j the determinant of the sample dispersion matrix of the j^{th} set of variates and by $D(j)$ the determinant of the dispersion matrix of the first j sets taken together. The Wilks' statistic used for testing the independence of the q sets is given by

(9) $$\Lambda = \frac{D(q)}{\prod_{j=1}^{q} D_j} = \lambda_q \prod_{j=2}^{q-1} \lambda_j,$$

where

$$\lambda_j = \frac{D(j)}{D_j D(j-1)}, \qquad j = 2, 3, \cdots q.$$

The region $W - w$ complementary to the critical region w is defined by

$$\Lambda \geq \text{a positive constant}$$

The statistic W is invariant to linear transformations within each set of variates. The distribution (8) may therefore without loss of generality be written in the form

(10) $$\prod_{j=1}^{q} \left[\prod_{r=l_{j-1}+1}^{l_j} \left\{(2\pi V_r^2)^{-(n/2)} \exp\left(-\frac{1}{2V_r^2} \sum_{i=1}^{n} (x_{ir} - \sum_{u=0}^{l_j-1} \beta_{ru} x_{iu})^2\right) \prod_{i,r} dx_{ir}\right\}\right].$$

Let B_j $(j = 2, 3, \cdots q)$ stand for the set of constants

$$\beta_{ru_j} \qquad \begin{aligned} r &= l_{j-1} + 1, l_{j-1} + 2, \cdots l_j, \\ u &= 1, 2, \cdots l_{j-1}, \end{aligned}$$

and let

$$B_j = 0 \tag{11}$$

imply the vanishing of all the constants of the set B_j. The q sets of variates will be independent if (11) holds for all values of j from 2 to q. Denote by $I(B_q, B_{q-1}, \cdots, B_2)$ the integral of (10) over the region $W - w$. Integrating first over the sub-region of $W - w$ for which

$$\prod_{j=2}^{q-1} \lambda_j$$

has a given value and using the result of section 4, it follows that

$$I(B_q, B_{q-1}, \cdots B_2) \leq I(0, B_{q-1}, \cdots B_2).$$

Also if $B_q = 0$, λ_q is distributed independently of λ_{q-1}. Hence starting with $B_q = 0$ in (10) and integrating for λ_{q-1} first, we obtain

$$I(0, B_{q-1}, B_{q-2}, \cdots B_2) \leq I(0, 0, B_{q-2}, \cdots B_2).$$

Thus finally, $I(B_q, B_{q-1}, \cdots B_2) \leq I(0, 0, \cdots 0)$, which proves the completely unbiassed character of the Wilks criterion.

REFERENCES

[1] J. E. DALY, "On the unbiased character of likelihood ratio tests for independence in normal system," *Annals of Math. Stat.*, Vol. 11 (1940), p. 1.

[2] P. C. TANG, "The power function of analysis of variance tests with tables and illustrations of their use," *Stat. Res. Mem.*, Vol. 2 (1938), p. 126.

[3] R. D. NARAIN, "A new approach to sampling distribution of the multivariate normal theory. Part I," *Jour. Ind. Soc. Agr. Stat.*, Vol. 1 (1948), p. 59.

[4] R. D. NARAIN, "A new approach to sampling distributions of the multivariate normal theory. Part II," *Jour. Ind. Soc. Agr. Stat.*, Vol. 1 (1948), p. 137.

ON THE DISTRIBUTION OF THE TWO CLOSEST AMONG A SET OF THREE OBSERVATIONS[1]

BY G. R. SETH

Iowa State College

1. Introduction. In this note we obtain the joint distribution of the two closest observation x', x'' $(x' < x'')$ of the set x_1, x_2, x_3 $(x_1 \leq x_2 \leq x_3)$ when the distribution of x_1, x_2, x_3 is given or can be obtained.[2] We will assume that in general the density function is given by $f(x_1, x_2, x_3)$ and that it is continuous in the

[1] The results in this paper were presented at a meeting of the Institute of Mathematical Statistics in Madison, Wisconsin, September 9, 1948.

[2] The author's attention was drawn to this problem while visiting the National Bureau of Standards in the Spring of 1948, by Mr. Julius Lieblein of the Statistical Engineering

variables involved We also find the distributions of certain statistics depending on x' and x''. We will denote the density and the cumulative distribution function of a normal variate with mean zero and unit variance by $\phi(x)$ and $G(x)$

2. Distribution of the two closest. Let x', x'' be the two closest among the set of x_1, x_2, x_3 $(x_1 \leq x_2 \leq x_3)$ Let $P(S_1, S_2, \cdots, S_k)$ denote the probability that the events $S_1, S_2, \cdots, S_k$ occur. Let us consider $P(x' < s, x'' < t)$, for $t < s$. For $s < t$, it reduces to $P(x'' < t)$ i.e. the marginal cumulative distribution of x''
Now

$$P(x' < s, x'' < t) = P(x_1 < s, x_2 < t, x_2 - x_1 < x_3 - x_2) \tag{1}$$
$$+ P(x_2 < s, x_3 < t, x_2 - x_1 < x_3 - x_2).$$

The equalities, here as well as elsewhere, are omitted as the variables admit continuous distributions. Let the first and second terms on the right side in (1) be denoted by $P(A)$ and $P(B)$ respectively, where A, B denote the events in the respective brackets The event B can be further split up into more elementary events whose probabilities can be easily found (B) can be seen to be equivalent to

$$\left(x_1 < 2s - t, \quad x_1 < x_2 < \frac{x_1 + t}{2}, \quad x_2 < x_3 < 2x_2 - x_1\right)$$
$$+ (2s - t < x_1 < s, \quad x_1 < x_2 < s, \quad x_2 < x_3 < 2x_2 - x_1)$$
$$+ \left(x_1 < 2s - t, \quad \frac{x_1 + t}{2} < x_2 < s, \quad x_2 < x_3 < t\right).$$

We may write (1) in the form of integrals and differentiating under the integral sign with respect to t and s we obtain

$$\frac{\partial^2 P}{\partial s \partial t} = \int_{2t-s}^{\infty} f(s, t, x_3)\, dx_3 + \int_{-\infty}^{2s-t} f(x_1, s, t)\, dx_1 \tag{2}$$

The right hand side of (2) gives the density function of x', x'' at $x' = s$, $x'' = t$. Let $f_{ij}(x_i, x_j)$ be the density function of x_i and x_j $(i > j = 1, 2, 3)$ Then the density function $p(x', x'')$ of x' and x'' can be put into the form

$$p(x', x'') = f_{12}(x', x'')\,[1 - F_3(2x'' - x' \mid x_1 = x', x_2 = x'')] \tag{3}$$
$$+ f_{23}(x', x'')\,[F_1(2x' - x'' \mid x_2 = x', x_3 = x'')],$$

where $F_i(x_i \mid x_j = l, x_k = m)$ represents the cumulative distribution function of the conditional density function of x_i when x_j and x_k are fixed at the values l and m respectively If, before ordering, the three observations are independent

Laboratory He understands that Mr Lieblein has in preparation for submission to the *Journal of Research of the National Bureau of Standards* a paper giving intensive consideration to the closest pair and other aspects of samples of three observations

and from the same population having the density function $f(x)$, then (3) with the help of

$$f(x_1, x_2, x_3) = 6f(x_1) f(x_2) f(x_3)$$

reduces to

$$p(x', x'') = 6f(x')f(x'')[1 - F(2x'' - x') + F(2x' - x'')] \tag{4}$$

where $F(x) = \int_{-\infty}^{x} f(x)\, dx$.

3. Joint distribution of $(x'' - x')$ and $(x'' - x')/(x_3 - x_1)$. Let $F_1(s, t)$ denote the cumulative distribution function of $u = x'' - x'$ and $w = \dfrac{x'' - x'}{x_3 - x_1}$. Then

$$F_1(s, t) = P\left[x'' - x' < s, \frac{x'' - x'}{x_3 - x_1} < t\right]. \tag{5}$$

The range for u is $(0, \infty)$ and w varies between 0 and $\frac{1}{2}$, and thus we limit ourselves to s varying from 0 to ∞, and t varying in $(0, \frac{1}{2})$.

After some manipulation of the probability statement and differentiating with respect to s and t under the integral sign, in a manner similar to that of the previous section, we obtain the joint density function of u and v, given by

$$\begin{aligned} \frac{\partial^2 F_1(s, t)}{\partial s \partial t} &= \frac{s}{t^2}\left[\int_{-\infty}^{\infty} f\left(x_1, x_1 + s, x_1 + \frac{s}{t}\right) dx_1 \right. \\ &\qquad \left. + \int_{-\infty}^{\infty} f\left(x_1, x_1 + \frac{s(1-t)}{t}, x_1 + \frac{s}{t}\right) dx_1\right] \\ &= f_1(s, t) \quad \text{(say)}. \end{aligned} \tag{6}$$

4. Applications to normal distributions. Let $f(x)$ in (4) be the density function of a normal distribution with mean θ and variance unity, then (6) reduces to

$$f_1(u, w) = \frac{2\sqrt{3}u}{\pi w^2} e^{-u^2 \frac{(1 - w + w^2)}{3w^2}}. \tag{7}$$

Further the marginal density of u and w will be given by

$$p(u) = 6\sqrt{2}\phi\left(\frac{u}{\sqrt{2}}\right)\left[1 - G\left(\frac{\sqrt{3}u}{\sqrt{2}}\right)\right], \tag{8}$$

$$p(w) = \frac{3\sqrt{3}}{\pi} \frac{1}{1 - w + w^2}, \qquad 0 < w < \tfrac{1}{2}, \quad \text{respectively} \tag{9}$$

The distribution of w has been obtained by J. Lieblein in an unpublished paper

From (2) we can also obtain the joint density function of $u = x'' - x'$ and

$v = \dfrac{x' + x''}{2}$. When we integrate this joint density function with respect to u, we obtain the density function of $v = \dfrac{x' + x''}{2}$ as given by

$$
\begin{aligned}
p(v) = 6\sqrt{2}\phi[\sqrt{2}(v - \theta)]\Big[1 + G\Big(\frac{\sqrt{2}(v - \theta)}{\sqrt{11}}\Big)& \\
- 2\int_0^\infty \phi(x) G\left(\frac{3x}{\sqrt{2}} + v - \theta\right) dx\Big].&
\end{aligned}
\tag{16}
$$

The mean and the variance of the distribution of v are given by θ and $\frac{1}{2} + \dfrac{\sqrt{3}}{4\pi}$ respectively.

It may be remarked that if there is a suspicion that one of the extreme observations in a sample of three does not belong to the normal population under consideration, then the median of the sample is a better estimate than the average of the two closest. The efficiency of the latter compared to that of the former is about 70%, for the variance of the median in this case is given by $1 + \dfrac{\sqrt{3}}{\pi}$ compared to $\frac{1}{2} + \dfrac{\sqrt{3}}{4\pi}$ of v, the average of the two closest. The efficiency is here defined as the ratio of the variances for the two estimates.

ERRATA

By W. Feller

Cornell University

The author regrets the following inconsequential, but very disturbing, slips in his paper "On the Kolmogorov-Smirnov limit theorems for empirical distributions" (*Annals of Math Stat*, Vol. 19 (1948), pp. 177–189):

(1) In equation (1 4) on p 178, the exponent $-\nu^2 z^2$ should be replaced by $-2\nu^2 z^2$. The same copying error occurs in the description of Smirnov's table on p. 279. The proof is correct as it stands.

(2) In the formulation of the *continuity-theorem* on p. 180 it is claimed that $u_k \to f(t)$ whereas in reality the continuity theorem permits only the conclusion that

$$
\delta \sum_{r=1}^{k} u_r \to \int_0^t f(x)\, dx. \tag{*}
$$

This slip in formulation in no way affects the proofs since only (*) is used. (The assertion that the step functions $\{\xi_k\}$ converge pointwise is not based on a

second application of the continuity theorem, but on the obvious fact that(*) implies

$$\delta \sum_{r=1}^{k} q_r u_r \to \int_0^t q(x) f(x)\, dx,$$

where the step function $\{q_r\}$ converges uniformly to a continuous monotonic $q(x)$).

The following corrections apply to the paper, "On the normal approximation to the binomial distribution" (*Annals of Math. Stat.*, Vol. 16, (1945), pp 319–329).

(1) Equation (27) gives two variants of an estimate for the error ρ. The second should simply restate the first one in terms of the variable x, in other words, the expression $(p^3 + q^3)$ in the second line of (27) should be replaced by $p^3(1 - px/\sigma)^{-3} + q^3(1 + qx/\sigma)^3$.

(2) The estimate $\rho < \sigma^{-6}/300$ given in (28) is not valid over the entire range for which it is claimed. However, the further theory depends only on the fact that $\rho = O(\sigma^{-4})$, and the estimate $\rho < \sigma^{-6}/30$ is both correct and sufficient for our purposes. (Actually, no changes whatever are required in the proofs, since (28) is used explicitly only for a range where it is correct as stated).

(3) On p. 324 it is stated that under the conditions of the main theorem (p. 325) $k \geq 4$, $n - k \geq 4$, whereas in reality the value 3 can occur in extreme cases. Fortunately, the assertion is not used anywhere in the proof, and the error ρ is negligible in all cases.

Accordingly, no changes are required either in the formulation or the proof of the theorems I am indebted to Dr W. Hoeffding for calling my attention to the slips.

(4) The first minus sign in footnote 5 should be an equality sign and the second minus in (70) a plus.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Chapel Hill meeting of the Institute, March 17–18, 1950)

1. A Method of Estimating the Parameters of an Autoregressive Time Series.

S. G. Ghurye, University of North Carolina.

The general autoregressive process of the second order is defined by the equations

$$x_t = X_t + \eta_t ,$$

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = \epsilon_t ,$$

where x_t is the value actually observed at time t, X_t the corresponding theoretical value, ϵ_t the disturbance and η_t the superposed variation. The estimates of α_1 , α_2 given by Yule's method are biased and inconsistent if η_t is not identically zero, the permanent bias being a function of the unknown variance of η_t The present paper proposes a method of estimation

which is unaffected by the presence of η_t, and seems to be better than any other known method; and this conjecture is supported by the results of application to observational and artificial series In this method the estimates a_1, a_2 are obtained by minimizing

$$\sum_{k=3}^{n} \frac{1}{(N-k-2)} \left\{ \sum_{t=3}^{N-k} (x_t + a_1 x_{t-1} + a_2 x_{t-2})(x_{t+k} + a_1 x_{t+k-1} + a_2 x_{t+k-2}) \right\}^2,$$

where n is some number small in comparison with N (which is the number of observations). In the above expression the usual approximation of substituting $(N - k - 2)r_k$ for $\sum_{t=3}^{N-k} x_t x_{t+k}$ may be made for computational convenience The method has been used for fitting autoregressive processes to the series of annual averages of Wolfer's sunspot numbers and that of Myrdal's Swedish cost of living index numbers The method is applicable to higher order processes

2. Most Powerful Rank Order Tests. (Preliminary Report). Wassily Hoeffding, University of North Carolina

Let $X_{11}, \cdots, X_{1n_1}, \cdots, X_{k1}, \cdots, X_{kn_k}$ be random variables with a joint probability function $P(S)$ and let $P\{X_{ig} = X_{ih}\} = 0$ if $g \neq h$ $(i = 1, \cdots, k)$. Let H_0 be a hypothesis which implies that $P(S)$ is invariant under all permutations of $X_{i1}, \cdots, X_{in_i}$ $(i = 1, \cdots, k)$ Let r_{ij} $(j = 1, \cdots, n_i)$ be the ranks of $X_{i1}, \cdots, X_{in_i}$. Under H_0 the $M = \Pi n_i!$ rank permutations $R = (r_{11}, \cdots, r_{1n_1}, \cdots, r_{k1}, \cdots, r_{kn_k})$ have the same probability $P(R) = M^{-1}$. A test which depends only on the permutations R is called a rank order test (R.O.T.). A R.O.T. of size m/M which is most powerful (M.P.) against a simple alternative, $P_1(S)$, is determined by m permutations R for which $P_1(R)$ takes on its m largest values.

For example, let the pairs $(X_1, Y_1, \cdots, (X_n, Y_n)$ be independent and identically distributed. Let H_0 state that X_i, Y_i are independent, and let $H_1(\rho)$ be the hypothesis that X_i, Y_i have a bivariate normal distribution with correlation ρ. We may assume that $X_1 < \cdots < X_n$ and consider the ranks r_i of the Y's only A R.O.T. which is uniformly M P. against all $H_1(\rho)$ with $\rho > 0$ does not exist except for small n The M P R O T against small $\rho > 0$ is determined by the largest values of $\sum_{i=1}^{n} (EZ_i)(EZ_{r_i})$, where EZ_i is the expectation of the i-th order statistic in a sample of n from a standard normal distribuion The M P. unbiased R.O.T against small values of $|\rho|$is based on the statistic $\sum_i \sum_j (EZ_iZ_j)(EZ_{r_i}Z_{r_j})$ The M.P. R.O T against ρ close to 1 is obtained by expanding the probability of $(r_1, \cdots, r_n)$ in powers of $\{(1-\rho)/(1+\rho)\}^{1/2}$

3. The Comparison of Percentages in Matched Samples. William G. Cochran, Johns Hopkins University.

In this paper the familiar χ^2 test for comparing the percentages of successes in a number of independent samples is extended to the situation in which each member of any sample is matched in some way with a member of every other sample This problem has been encountered in the fields of psychology, pharmacology, bacteriology, and sample survey design A solution has been given by McNemar (1949) when there are only two samples

In the more general case, the data are arranged in a two-way table with r rows and c columns, in which each column represents a sample and each row a matched group The test criterion proposed is

$$Q = \frac{c(c-1)\Sigma(T_j - \bar{T})^2}{c(\Sigma u_i) - (\Sigma u_i^2)},$$

where T_j is the total number of successes in the j^{th} sample and u_i the total number of successes in the i^{th} row If the true probability of success is the same in all samples, the limit-

ing distribution of Q, when the number of rows is large, is the χ^2 distribution with $(c - 1)$ degrees of freedom. The relation between this test and the ordinary χ^2 test, valid when samples are independent, is discussed.

In small samples the exact distribution of Q can be constructed by regarding the row totals as fixed, and by assuming that on the null hypothesis every column is equally likely to obtain one of the successes in a row. This exact distribution is worked out for eight examples in order to test the accuracy of the χ^2 approximation to the distribution of Q in small samples. The number of samples ranged from $c = 3$ to $c = 5$. The average error in the estimation of a significance probability was about 14 per cent in the neighborhood of the 5 per cent level and about 21 per cent in the neighborhood of the 1 per cent level. Correction for continuity did not improve the accuracy of the approximation, although it is recommended when there are only two samples. Another approximation, obtained by scoring each success as "1" and each failure as "0" and performing an analysis of variance on the data, was also investigated. The F-test, corrected for continuity, performed about as well as the χ^2 approximation (uncorrected), but is slightly more laborious.

The problem of subdividing χ^2 into components for more detailed tests is briefly discussed.

4. A Method of Estimating Components of Variance in Disproportionate Numbers. H. L. Lucas, North Carolina State College.

By including sufficient effects in the forward solution of the Abbreviated Doolittle method, components of variance may be estimated from disproportionate data. The procedure is very systematic, and thus, is adaptable to routine computational work. The computations will be described, and the utility of the method briefly discussed.

5. On the Theory of Unbiased Tests of Simple Statistical Hypotheses Specifying the Values of Two Parameters. (Preliminary Report). Stanley L. Isaacson, Columbia University.

In the Neyman-Pearson theory of testing simple hypotheses, in the one-parameter case, a locally best unbiased region is called "type A." It is obtained by maximizing the curvature of the power curve at the point $\theta = \theta_0$ specified by the hypothesis, subject to the conditions of size and unbiasedness. For the two-parameter case, Neyman and Pearson considered "type C" regions (*Stat. Res. Mem.*, vol. 2 (1938), p. 36). The definition of these regions requires one to choose in advance a family of ellipses of constant power in an infinitesimal neighborhood of the point $(\theta_1, \theta_2) = (\theta_1^0, \theta_2^0)$ specified by the hypothesis. The natural generalization of a "type A" region is a "type D" region, which maximizes the Gaussian curvature of the power surface at (θ_1^0, θ_2^0), subject to the conditions of size and unbiasedness. This definition does not require one to choose a family of ellipses in advance. This approach leads to a new problem in the calculus of variations. A sufficient condition is obtained which plays the role of the Neyman-Pearson fundamental lemma in the "type A" case. An illustrative example is given. (Prepared under sponsorship of the Office of Naval Research.)

6. A Note on Orthogonal Arrays. Raj Chandra Bose, University of North Carolina.

Consider a matrix $A = (a_{ij})$ with N rows and m columns, each element a_{ij} standing for one of the s integers $0, 1, 2, \cdots, s - 1$. Let us take the partial matrix obtained by choosing any $t \leq m$ columns of A. Each row now consists of an ordered t-plet of numbers, and each

element has one of s possible values, there are s^t possible t-plets. The matrix A may be called an orthogonal array (N, m, s, t) of size N, m constraints, s levels and strength t, if by choosing any t columns whatsoever every possible t-plet occurs the same number of times. Clearly $N = \lambda s^t$ where λ is an integer. Such arrays have been considered by Rao and are useful for various experimental designs. The existence of an orthogonal array $(s^2, M, s, 2)$ is equivalent to the existence of a set of orthogonal Latin squares of side s and m constraints (i.e., the number of Latin squares in the set is $m - 2$). The fundamental question that can be asked regarding orthogonal arrays is the following: What is the maximum number of constraints for an orthogonal array, given N, s and t? Denote this number by $f(N, s, t)$, then from known properties of Latin squares $f(s^2, s, 2) = s + 1$, if s is a prime or a prime power, and a theorem by Mann states that $f(s^2, s, 2) = r + 1$, if $s = p_1^{n_1} \cdots p_k^{n_k}$, where $p_1, \cdots, p_k$ are different primes, and r is the minimum of $p_1^{n_1}, p_2^{n_2}, \cdots, p_k^{n_k}$. The following generalisation of Mann's theorem is proved in this note

$$f(N_1 N_2 \cdots N_k, s_1 s_2 \cdots s_k, t) = \operatorname{Min}\{f(N_1, s_1, t), f(N_2, s_2, t), \cdots, f(N_k, s_k, t)\}.$$

7. Transformations Related to the Angular and the Square Root. MURRAY F. FREEMAN AND JOHN W. TUKEY, Princeton University

The use of transformations to stabilize the variance of binomial or Poisson data is familiar (Anscombe, Bartlett, Curtiss, Eisenhart). The comparison of transformed binomial or Poisson data with percentage points of the normal distribution to make approximate significance tests or to set approximate confidence intervals is less familiar. Mosteller and Tukey have recently made a graphical application of a transformation related to the square-root transformation for such purposes, where the use of "binomial probability paper" avoids all computation. We report here on an empirical study of a number of approximations, some intended for significance and confidence work, and others for variance stabilization. (Prepared in connection with research sponsored by the Office of Naval Research.)

8. Standard Inverse Matrices for Fitting Polynomials. F. J. VERLINDEN, North Carolina State College.

For fitting polynomials of the type, $y = b_0 x^0 + b_1 x + b_2 x^2 + \cdots + b_m x^m$, with the x's equally spaced, published tables of orthogonal polynomials may be used. This procedure does not yield the b's directly, nor their variances or covariances, although such may be obtained by proper computations which are moderately tedious. In some types of statistical work, the b's and their variances and covariances may be desired. These may of course be obtained directly by the method of least squares but the computational work is prodigious relative to that for the orthogonal polynomial approach. When the x's are equally spaced the elements of the variance-covariance matrix may be put in the simple form of sums of powers (including the zero power) of successive integers from zero to n (n equals one less than the number of observations). The elements of the inverses of matrices of this type have been worked out algebraically in terms of n for polynomials up to and including the quintic ($m = 5$). With these standard inverse matrices, the b's and their variances and covariances may quickly be obtained once the elements are evaluated numerically. These elements have been evaluated numerically up to $n = 20$.

9. Mathematical Models in Biology. J. A. RAFFERTY, Department of Biometrics, School of Aviation Medicine, Randolph Field, Texas.

From the point of view of a bio-medical research administrator, mathematical models

will assume a greater role in biological research than heretofore. In anticipation of this trend, certain philosophical implications of models in biological theory and scientific theory in history are examined. A hierarchy of abstraction-levels in biology is delineated, and the role of mathematical models at these levels is illustrated by examples from the literature. Proposals are made for a concentration of mathematical effort on certain important biological problems. Remarks are made on the capabilities and limitations of models in biology.

10. Small Sample Performance of Biological Statistics. Irwin Bross, Johns Hopkins University.

In this paper the dilution method for estimating bacterial density is investigated by an exact small sample method and also by an approximate one. Methodologies and design of experiments are compared for various small sample cases.

11. Methodology in the Study of Physical Measurements of School Children. B. G. Greenberg and A. Hughes Bryan, University of North Carolina.

In a series of investigations to determine by small-sampling technique what physical differences, if any, occur between children of differing socio-economic backgrounds, several problems of methodology arose. A pilot study was undertaken to assure maximum efficiency at each step. This paper reports some of these results. It was found that the children could remain dressed (with the exception of boys' bi-iliac measurement) without changing the magnitude of the differences. The pilot study enabled us to decide how many observers to use, and how much duplication of measurements by them was necessary. Minimum sample sizes were estimated to indicate physical differences of predetermined magnitudes. It was found that the age grouping 96–143 months was optimal from the standpoint of indicating physical differences between children of differing socio-economic levels. Boys and girls in the upper socio-economic levels were both taller and heavier for their age in this age group. There were no weight differences, however, when weight was adjusted for age and height. Measurement of the bi-iliac and transverse chest diameter provided little additional information on physical differences. The calf circumference, an indicator of muscle mass and subcutaneous fat, is suggested as being a sensitive supplementary index to indicate physical differences when age and height are adjusted.

12 Tetrad Analysis in Yeast. A. S. Householder, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

In neurospora all four products of meiosis are recovered in the four spores of an ascus. In crosses $AB \times ab$ the asci are of three types, designated I, II or III according as all four, none, or two spores resemble parents. Frequencies of these types, P, P′ and P″ are the observables. If there were no exchange P'' would be zero; and one should have $P' = 0$ or $\frac{1}{2}$ according to whether the loci were on the same or different chromosomes.

Assuming only that no exchange occurs between sister chromatids and neglecting chromatid interference, one can calculate without further assumptions a frequency P'' of exchanges between a single locus and its centromere from data on three or more genes taken in pairs by equations

$$s_{ij} = s_{0i}s_{0j}\,, \qquad P'' = 2(1 - s)/3,$$

where the subscript 0 refers to a centromere. Lindegren makes such calculations from his own data, by taking groups of three, but makes no effort to reconcile discrepancies. Neyman's modified chi-square, however, permits combining all observations in a set of equa-

tions that yields easily to rapidly converging iterative solution. The equations are

$$2s_i \sum_{j \neq i} s_j^2 (n_{ij} + n'_{ij})^2 (n_{ij}^{-1} + n'^{-1}_{ij}) = \sum_{j \neq i} s_j (n_{ij} + n'_{ij})^2 (2n_{ij}^{-1} - n'^{-1}_{ij}),$$

where n_{ij} is the number in class I and II combined for the loci i and j, n'_{ij} the number class in III, and only those pairs (i, j) are included which are found to be independent

The argument of A R. G. Owen (*Proc. Roy. Soc*, *Ser B*, Vol 136 (1949) pp 67–94) can be paraphrased for the present case and a suitable generating function $P(\lambda, u)$ is being sought providing a metric. The specific one proposed by Owen is ruled out since $s = P(-\frac{1}{2}, u)$ takes on a negative value for one locus, which is not possible with Owen's function.

13. Contribution to the Probabilistic Theory of Neural Nets. I. Randomization of Refractory Periods and of Stimulus Intervals. ANATOL RAPOPORT, University of Chicago.

Aggregates of neurons are considered in which the frequency of occurrence of neurons with a specified value of the refractory period follows certain probability distributions Input-output functions are derived from such aggregates In particular, if input and output intensities are defined in terms of stimulus frequencies and firing frequencies per neuron respectively, it is shown that a rectangular distribution of refractory periods leads to a logarithmic input-output curve. If input and output are defined in terms of the total number of stimuli and firings in the aggregate, it is shown how the "mobilization" picture leads to the logarithmic input-output curve.

By randomizing the intervals between stimuli received by a single neuron and by introducing an inhibitory neuron a very simple "filter net" can be constructed whose output will be sensitive to a particular range of the input, and this range can be made arbitrarily small.

14. Theoretical and Experimental Aspects in the Removal of Air-Borne Matter by the Human Respiratory Tract. H. D. LANDAHL, University of Chicago

The principal factors governing the fate of a particle in the respiratory tract are impaction due to inertia, settling due to gravity and Brownian movements. For a given respiratory pattern, it is possible to calculate the probable fate of a particle from a knowledge of the geometry of the passages. These calculations have been carried out in such a manner as to obtain the theoretical amounts of material deposited in various regions of the lungs as well as the relative amounts in various fractions of the expired air Similarly, it is possible to estimate the probable fate of a particle which passes through the nasal passages. Experiments have been carried out to verify a number of these predictions On the whole, the agreement, as illustrated in the slides, is fairly satisfactory when one considers the complexity of the calculations.

15. An Application of Biometrics to Zoological Classification. F. M. WADLEY, Navy Department, Washington, D. C.

Statistical problems in taxonomy are discussed, attention must be paid to variation of individuals as well as of group means. Covariance analysis and the discriminant function technique are applied to multiple measurements in groups of molluscan fossils

16. The Analysis of Hemotological Effects of Chronic Low-Level Radiation. JACK MOSHMAN, United States Atomic Energy Commission, Oak Ridge, Tennessee.

Several methods are investigated for analyzing the possible effects of chronic low-level irradiation upon the employees of the operating contractors of the US AEC. The effects investigated are those on the red blood count, hemoglobin, white blood count, lymphocytes and neutrophils. The analysis includes measurements of significant differences among individuals, geographic sites and the exploration of various indices of exposure to radiation. A non-parametric determination of trend values for individuals which may be applied to mass data is considered

17. Statistical Problems in Psychological Testing. EDWARD E. CURETON, University of Tennessee

Though great progress has been made in mathematical statistics in recent years, a number of the major statistical problems encountered in the development and use of psychological tests remain unsolved Some of these problems are outlined, with particular reference to the mathematical models and assumptions implied by psychological theory, by the nature of the experimental data, and by the conditions under which the results and findings are to be applied.

18. Accuracy of a Linear Prediction Equation in a New Sample. GEORGE E. NICHOLSON, JR., University of North Carolina.

The problem considered is as follows. Given two samples S_1 and S_2 of N_1 and N_2 observations on a $p + 1$ character random variable $(y, x_1 \cdots x_p)$. Let Y_1 and Y_2 be the linear regression equation computed by the method of least squares from each sample The effect of using Y_1 to predict the y's in S_2 is considered. The ratio k $\dfrac{S(y_2 - Y_1)^2}{S(y_2 - Y_2)}$ s used as a measure of the predicting efficiency of Y_1 in S_2 relative to Y_2 when the X_i are fixed for the usual regression model The general multivariate case is also considered

19. Independence of Quadratic Forms in Normally Correlated Variables. YUKIYOSI KAWADA, Tokyo University of Literature and Science, Tokyo, Japan.

An extension is given of theorems of Craig, Hotelling and Matérn which includes the following theorem, proved by a new method If two quadratic forms Q_1 , Q_2 in normally and independently distributed variates with zero means and unit variances satisfy the four conditions $E(Q_1^iQ_2^j) = E(Q_1^i)E(Q_2^j)$, for $i, j = 1, 2$, then the product of the matrices of the two forms in either order is zero

20. Bounds on the Distribution of Chi-square. S. A. VORA, University of North Carolina.

Let

$$\chi^2 = \sum_{j=1}^{k} (v_j - np_j)^2/np_j, \qquad \chi'^2 = \sum_{j=1}^{k} (v_j + \tfrac{1}{2} - Np_j)^2/Np_j,$$

where $v_j \geq 0, \sum_1^k v_j = n, p_j > 0, \sum_1^k p_j = 1$ and $N = n + k/2$. Bounds on the multinomial probability T in terms of χ'^2 are obtained A triangular transformation of

$$x_i = (v_i + \tfrac{1}{2} - Np_i)/\{Np_i(1 - p_i)\}^{1/2} \qquad (i = 1, \cdots, k - 1),$$

to y_i is applied so that

$$d\ \chi'^2 = \sum_{i=1}^{j-1} y_i^2,$$

where d is determined later by equating the coefficients of χ'^2 Certain rectangles $r(v)$ with $(y_1, \cdots, y_{k-1})$ as a mid-point are non-overlapping and cover the entire space R_{k-1} for $v_i = 0, \pm 1, \pm 2, \cdots$. If $\chi'^2 \leq c$, then bounds on T in terms of the integral of the $(k-1)$ dimensional normal frequency function over the rectangle $r(v)$ are obtained Prob $\{\chi'^2 \leq c\}$ is the sum of T over $\chi'^2 \leq c$, so the integral over the sum of rectangles whose mid-points lie within the hypersphere $\chi'^2 \leq c$ is considered. Two hyperspheres, one which contains the sum of those rectangles, and one which is contained in it are used for the bounds, giving

$$\lambda_2\ F_{k-1}(c_2) \leq \text{Prob}\ \{\chi'^2 \leq c\} \leq \lambda_1 \cdot F_{k-1}(c_1),$$

where $F_{k-1}(c)$ is a chi-square distribution function with $(k-1)$ degrees of freedom and $\lambda_1, \lambda_2, c_1, c_2$ are functions of c, n, k and $p_1, \cdots, p_k$. As $n \to \infty$, both bounds tend to $F_{k-1}(c)$. Bounds of the same form are obtained for Prob $\{\chi^2 \leq C\}$ Closer bounds for Prob. $\{\chi^2 \leq C\}$ are given in terms of a non-central chi-square distribution

21. Estimation of Genetic Parameters. C. R. HENDERSON, Cornell University.

Many applications of genetics and statistics to the improvement of plants and animals deal with experimental data for which the underlying model is assumed to be

$$y_\alpha = \sum_{i=1}^{p} b_i x_{i\alpha} + \sum_{i=1}^{q} u_i z_{i\alpha} + e_\alpha,$$

where b_i are unknown fixed parameters, $x_{i\alpha}$ and $z_{i\alpha}$ are observable parameters, the u_i are a random sample from a multivariate normal distribution with means zero and covariance matrix $\| \sigma_{ij} \|$, and the e_α are normally and independently distributed with means zero and variances σ_α^2. If $\sigma_{ij} = 0$ when $i \neq j$ and if $\sigma_\alpha^2 = \sigma_e^2$, the model is the one usually assumed when components of variance are estimated

Three different estimation problems are involved, (1) estimation of b_i under the assumptions of the model, (2) estimation of u_i and (3) estimation of σ_{ij}. The first two problems are not solved satisfactorily by the least squares procedure in which the u_i are regarded as fixed, but the maximum likelihood solution does lead to a satisfactory estimation procedure.

Assuming that the σ_{ij} and σ_α^2 are known, the joint maximum likelihood estimates of b_i and u_i are the solution to the set of linear equations

$$\sum_{i=1}^{p} b_i \left(\sum_\alpha x_{h\alpha} x_{i\alpha}/\sigma_\alpha^2\right) + \sum_{i=1}^{q} u_i \left(\sum_\alpha x_{h\alpha} z_{i\alpha}/\sigma_\alpha^2\right) = \sum_\alpha x_{h\alpha} y_\alpha/\sigma_\alpha^2, \qquad h = 1, \cdots, p,$$

$$\sum_{i=1}^{p} b_i \left(\sum_\alpha x_{i\alpha} z_{h\alpha}/\sigma_\alpha^2\right) + \sum_{i=1}^{q} u_i\left(\sigma^{ih} + \sum_\alpha z_{i\alpha} z_{h\alpha}/\sigma_\alpha^2\right) = \sum_\alpha z_{h\alpha} y_\alpha/\sigma_\alpha^2, \qquad h = 1, \cdots, q.$$

Some important applications of this estimation procedure to genetic studies are described and certain computational short-cuts are suggested

The problem of estimating σ_{ij} has not been solved satisfactory although under certain quite general assumptions the equations for the joint estimation of b_i, u_i, σ_{ij}, and σ_α^2 can easily be written The solution to the equations, however, is too difficult to make the procedure practical Nevertheless unbiased estimates of σ_{ij} can be obtained by equating to their expected values the differences between certain reductions in sums of squares computed by least squares and solving for the σ_{ij} In general, the expectation of the reduction due to $b_1, \cdots, b_p, u_1, \cdots, u_k (k \leq q)$ is $\sum \sum_{gh} d^{gh} E(Y_g Y_h)$, where d^{gh} are the elements

of the matrix which is the inverse of the $(p+k)^2$ matrix of coefficients and the Y_g are the right members of the least squares equations

22. Estimating the Mean and Standard Deviation of Normal Populations from Double Truncated Samples. A. C. COHEN, JR., University of Georgia.

The method of maximum likelihood is employed to obtain estimates of the mean and standard deviation of a normally distributed population from double truncated random samples Two cases are considered In the first, the number of missing variates is assumed to be unknown In the second, the number of missing (unmeasured) variates in each tail is known Variances for the estimates involved in each case are obtained from the maximum likelihood information matrices. A numerical example is given to illustrate the practical application of the estimating equations obtained for each of the two cases considered

23. Minimax Estimates of Location and Scale Parameters. GOPINATH KALLIANPUR, University of North Carolina.

If the joint fr f of the random variables $X_1, \cdots, X_N$ contains only a scale parameter and is of the form

$$\frac{1}{\alpha^N} p\left(\frac{x_1}{\alpha}, \cdots, \frac{x_N}{\alpha}\right),$$

then under mild restrictions the following theorem is proved:

THEOREM 1 *If the loss function is of the form* $W\left(\frac{\alpha-\bar{\alpha}}{\alpha}\right)$, *the best or minimax estimate* $\bar{\alpha}_0(x)$ *of* α *minimizes*

$$\int_0^\infty W\left(\frac{\alpha-\bar{\alpha}}{\alpha}\right)\frac{1}{\alpha^N} p\left(\frac{x_1}{\alpha}, \cdots, \frac{x_N}{\alpha}\right) d\alpha$$

w r t. $\bar{\alpha}$ *and further,*

$$\bar{\alpha}_0(\mu x_1, \cdots, \mu x_N) = \mu\bar{\alpha}_0(x_1, \cdots, x_N), \qquad \mu > 0$$

When both location and scale parameters are present and the joint fr f is of the form

$$\frac{1}{\alpha^N} p\left(\frac{x_1-\theta}{\alpha}, \cdots, \frac{x_N-\theta}{\alpha}\right),$$

(under conditions similar to those in Theorem 1) we obtain two results for the estimation of θ and α, respectively, one of which is

THEOREM 2 *If the loss function is of the form* $W\left(\frac{\theta-\bar{\theta}}{\alpha}\right)$, *the best estimate* $\bar{\theta}_0(x)$ *of* θ *minimizes*

$$\int_{-\infty}^\infty \int_0^\infty W\left(\frac{\theta-\bar{\theta}}{\alpha}\right)\frac{1}{\alpha^N} p\left(\frac{x_1-\theta}{\alpha}, \cdots, \frac{x_N-\theta}{\alpha}\right) d\theta\, d\alpha$$

$$\textit{and} \qquad \bar{\theta}_0\left(\frac{x_1+\lambda}{\mu}, \cdots, \frac{x_N+\lambda}{\mu}\right) = \frac{\bar{\theta}_0(x_1, \cdots, x_N)+\lambda}{\mu}$$

These theorems have been applied to derive minimax estimates in the case of standard distributions Finally, the problem of estimating the difference between the location parameters of two populations is briefly considered The results obtained in this paper are a continuation of the line of approach suggested in Theorem 5 of Wald's, "Contributions

to the Theory of Statistical Estimation and Testing Hypotheses" (*Annals of Math Stat*, Vol. 10 (1939), pp 299–225) (The present work was carried out under Office of Naval Research contract)

24. On Some Features of the Neyman-Pearson and the Wald Theories of Statistical Inference, Their Interrelations and Their Bearing on Some Usual Problems of Statistical Inference. S. N. Roy, University of North Carolina

With two alternative hypotheses H_1 and H_2 it is shown that (i) the most powerful test of H_1 with respect to H_2 is automatically an unbiased test in the sense that its power is never less than (and usually greater than) the level of significance α and (ii) there is also a least powerful test with its power not greater (usually less) than α This means that all tests have powers lying in between, which gives a complete picture of the possible family of tests and provides a basis for defining efficiency of tests.

With the first kind of error α is tied up a minimum second kind of error β (complementary to the maximum power P), and the level at which α is fixed depends upon some compromise between α and β. This intuitive approach is formalised by the introduction of loss functions related to and apriori probability weights for H_1 and H_2, thus leading to the first stage in the Wald treatment of dichotomy with two solutions in the observation space corresponding respectively to minimum and maximum total risks This is immediately generalised to the first stage in the Wald treatment of multichotomy with minimum and maximum total risk solutions. An important special case is discussed in which all the possible alternatives to a particular hypothesis are, by our test procedure, indistinguishable among themselves, thus effectively forming only one alternative to the hypothesis, which means a degenerate multichotomy The bearing of this on most powerful tests on an average under the Neyman-Pearson theory is also discussed

The problem of testing a composite hypothesis which is usually treated in terms of the Neyman-Pearson theory is posed and treated in terms of the (first stage) Wald theory and an indication is given of how these notions could be applied to the usual problems of univariate and multivariate analysis

25 Note on Uniformly Best Unbiased Estimates. R. C Davis, Naval Ordnance Test Station, Inyokern, California.

For the estimation in an absolutely continuous probability distribution of an unknown parameter which does not possess a sufficient statistic, it is shown that no unbiased estimate for the unknown parameter exists which attains minimum variance uniformly over a parameter set of arbitrary nature This result demonstrates the impossibility of obtaining a generalized sufficient statistic first proposed by Bhattacharyya Although not used in this note it is surmised that Barankin's powerful results on locally best unbiased estimates can be applied to yield further results in this direction.

26. Competitive Estimation. Herbert Robbins, University of North Carolina.

Let θ be a vector random variable with distribution function $G(\theta)$ and let x be a vector random variable whose frequency function $f(x; \theta)$ depends on θ Two statisticians, A and B, are required to estimate θ from the value of x If A's estimate is closer to θ he wins one dollar from B, and *vice versa*; in case of a tie no money changes hands It is shown that A should estimate θ by the function $a(x)$ = median of posterior distribution of θ given x, his expected gain will then be ≥ 0 whatever estimate B may use. If $G(\theta)$ is not known to A he should estimate it from the series of values of θ which have been observed in previous

trials. If these are not known, A should estimate $G(\theta)$ from the values of x which have previously occurred, how this may be done is discussed elsewhere (see Abstract 35)

From the point of view of the theory of games, when $G(\theta)$ is unknown we have a game in which the "rules" are unknown and must be successively estimated from past experience. Other examples arise whenever a game involves random devices whose probability distributions are not known to the players but must be inferred by statistical methods, in general from secondary variables which contain only part of the total information The role of statistical inference in such "long term" games is fundamental

27 The Effect of an Unknown 'Location Disturbance' on "Student's" t based on a Linear Regression Model. UTTAM CHAND, Boston University

Consider $y_1, \cdots, y_{N_1}, y_{N_1+1}, \cdots y_N$, a set of observations ordered in time If the y's are normally and independently distributed according to $N(\alpha + \beta(t - \bar{t}), \sigma^2)$ and we want to find out if the y's have changed with time, we usually employ a "Student's" t type of statistic with $N - 2$ degrees of freedom. If, as a consequence of the impact of a certain unknown political or economic change in the past on the y's, the y's actually constitute two independent, normal samples $y_1, \cdots y_{N_1}, y_{N_1+1}, \cdots y_N$ distributed according to $N(m_1, \sigma^2)$, $N(m_2, \sigma^2)$ respectively, a two-sample "Student's" t also based on $N - 2$ degrees of freedom would be the appropriate statistic to use for the hypothesis $m_1 = m_2$. If, in fact, the latter situation describes the correct state of affairs, and the statistician employs the "Student's" t based on the regression model, he commits an error The present paper investigates the nature of such an error in the light of the point of impact as determined by the magnitude of N_1 and the intensity of the impact as determined by the standardized 'distance' $\dfrac{m_2 - m_1}{\sigma\sqrt{\dfrac{1}{N_1} + \dfrac{1}{N - N_1}}}$ of this extraneous 'shock' on the ordered set of observations y

28 Corrections for Non-normality for the Two-sample t and the F Distributions Valid for High Significance Levels. RALPH A. BRADLEY, McGill University.

The effects of non-normality of the parent population on common tests of significance have long been of concern in the application of statistical methods to experimental data In this paper, the two-sample t-statistic is expressed as a simple multiple of the cotangent of an angle between two lines in a space of dimensionality one less than the total of the sample sizes, the F-statistic for k samples is expressed as a multiple of the cotangent of an angle between a line and a plane of $(k - 1)$ dimensions in a space, again, of dimensionality one less than the total of the sample sizes. The geometrical formulation is such as to suggest approximations to the distributions of these statistics valid for large values of the statistics, and these approximations are obtained. The approximations are shown to be exact in the special cases where the parent population is normal, and a method of evaluation of correction factors is given for a wide class of parent populations The approximation procedures are valid for the distributions under both null and non-null hypotheses

29 Some Tests Based on the Empirical Distribution Function. (Preliminary Report). JAMES F. HANNAN, University of North Carolina.

Let $X = (X_1, X_2, \cdots, X_n)$ be an independent sample of n where X_i has the continuous c d f. $F(x)$ Let $S_n(x)$ be the empirical distribution function Acceptance regions of

the type $\{X : S_n(x) \leq \phi(x) \text{ for all } x\}$ are considered for different specifications of ϕ and their probabilities evaluated The method of evaluation consists in identifying the regions with regions defined in terms of the order statistics of a sample of n from the uniform distribution on the interval (0, 1) The result obtained for $\phi(x) = F(x) + c/n$, $0 \leq c$, integral $\leq n$ is used to provide a direct proof of the Kolmogoroff result

$$\lim_{n \to \infty} P[n^{1/2} \sup_x (S_n(x) - F(x)) \leq z] = 1 - e^{-2z^2},$$

while that obtained for $\phi(x) = F(x) + t$, $0 \leq t \leq 1$, gives the exact c.d f. of the statistic $\sup_x (S_n(x) - F(x))$.

30. On a Generalization of the Behrens-Fisher Problem. (By Title). JOHN E. WALSH, Rand Corporation, Santa Monica, California.

Let $m + n$ independent observations be available where it is only known that a specified m of them are from continuous symmetrical populations with common median μ while the remaining n are from continuous symmetrical populations with common median ν. This is the generalization of the Behrens-Fisher problem investigated; some tests and confidence intervals for $\mu - \nu$ which are valid for the generalized situation are presented For definiteness, suppose that $n \leq m$ The procedure used is to subdivide the m observations (common median μ) into n groups of nearly equal size and form the mean of the observations for each group. Pair the n means with remaining n observations and subtract the value of each observation from the value of the mean with which it is paired. The resulting n values represent independent observations from populations with common median $\mu - \nu$ Tests and confidence intervals for $\mu - \nu$ are obtained by applying the results of "Applications of Some Significance Tests for the Median Which are Valid Under Very General Conditions" (*Jour. Amer. Stat. Assn.*, Vol 44 (1949), pp. 342–55) to these n values To measure the "information" lost by using the generalized tests when one actually has two independent samples from normal populations, power efficiencies are computed with respect to. (a) Scheffé's "best" t-test solution and (b) most powerful solution when ratio of variances is known. Case (a) yields an upper bound while case (b) furnishes a lower bound for the actual efficiency.

31. Construction of Partially Balanced Designs with two Accuracies. (By Title). S. S. SHRIKHANDE, University of North Carolina and Nagpur College, Nagpur, India.

Various methods of construction of partially balanced designs first introduced by Bose and Nair (*Sankhyā*, Vol 4 (1939), pp. 337–373) have been considered. Two of the methods given are generalisations of a difference theorem given by them. Another method is the inversion of an unreduced balanced incomplete block design with $k = 2$. Use has also been made of the existing balanced incomplete block design in another direction A number of designs can also be obtained by methods of finite geometries and especially by omitting a number of treatments and certain blocks from the complete lattice designs Use of curves and surfaces in finite geometries and the use of multifactorial designs given by Plackett and Burman (*Biometrika*, Vol. 33 (1946), pp 305–325) are also indicated

32. Designs for Two-way Elimination of Heterogeneity. (By Title). S. S SHRIKHANDE, University of North Carolina and Nagpur College, Nagpur, India.

Use has been made of the existing balanced and some partially balanced designs for two-

way elimination of heterogeneity with at most two accuracies. Particular cases of these designs were given by Youden (*Contributions from Boyce Thompson Institute*, Vol 9 (1937), pp. 317–326) and Bose and Kishen (*Science and Culture* (1939), pp. 136–137) The method depends upon interchanging the positions of various treatments in the different columns (blocks), if necessary, so as to satisfy certain conditions.

33. Designs for Animal Feeding Experiments. (By Title). S S. SHRIKHANDE, University of North Carolina and Nagpur College, Nagpur, India

In animal-feeding experiments change-over designs are generally preferable to continuous feeding experiments In change-over designs both the direct and carry-over treatment effects are important Use of balanced and partially balanced incomplete block designs toward this end has been considered

34. A Truncated Sequential Procedure for Interval Estimation, with Applications to the Poisson and Negative Binomial Distributions. (Preliminary Report). (By Title). D MARTIN SANDELIUS, University of Uppsala, Sweden, and University of Washington

Let $x, y_1, y_2, \cdots$ be a sequence of random variables defined in $(0, \infty)$, and let n be the smallest integer satisfying $\sum_{i=1}^{n+1} y_i > tx$, where $t > 0$ is a non-random quantity Define u_k either as $\sum_{i=1}^{k} y_i/x$ or as the smallest integer exceeding $\sum_{i=1}^{k} y_i/x$, $k = 1, 2, \cdots$. Given the distribution function $F(x, \theta)$ of x and, for any t, the conditional distribution of n with respect to x, the distribution of u_k is obtained The problem is to determine a confidence interval for θ with confidence coefficient $1 - \alpha$ on the basis of either an observation on u_k, if $u_k \leq t$, or an observation on n, if $n \leq k - 1$ The following procedure is proposed. If $u_k \leq t$, choose θ_{10} and θ_{11} according to a rule satisfying Prob $(\theta_{10} \leq \theta \leq \theta_{11}) \mid u_k \leq t) \geq 1 - \alpha$ If $n \leq k - 1$, choose θ_{20} and θ_{21} such that Prob $(\theta_{20} \leq \theta \leq \theta_{21} \mid n \leq k - 1) \geq 1 - \alpha$ For continuous u_k the following cases are discussed. a) $x = \theta$ with probability 1, and n has, for any t, a Poisson distribution with mean $t\theta$, b) x has a Gamma distribution with mean θ, and the conditional distribution of n with respect to x is, for any t, a Poisson distribution Both cases may, for instance, be applied to bacterial counting.

35. A Generalization of the Method of Maximum Likelihood: Estimating a Mixing Distribution. (Preliminary Report). (By Title). HERBERT ROBBINS, University of North Carolina

Let θ be a vector random variable with distribution function $G(\theta)$ belonging to some class $\mathfrak{G}$, let x be a vector random variable whose frequency function $f(x; \theta)$ depends on θ, and let $g^*(x) = \int f(x, \theta)\, dG(\theta)$ be the resulting frequency function of x From a sample $x_1, x_2, \cdots$ it is required to estimate $G(\theta)$ The generalized method of maximum likelihood consists in using the estimates $G_n(\theta; x_1, \cdots, x_n)$ in $\mathfrak{G}$ for which $\Pi\, g^*(x_i)$ is a maximum Under certain restrictions this method is consistent as $n \to \infty$.

Any consistent method of estimating the mixing distribution $G(\theta)$ from the sequence $x_1, x_2, \cdots$ yields a solution of parametric statistical decision problems in the following manner from *past* values $x_1, \cdots, x_{n-1}$ we estimate $G(\theta)$, and then use the corresponding Bayes solution of the decision problem to reach our decision for x_n, even though the value θ_n which produced x_n is different from those which produced $x_1, \cdots, x_{n-1}$ In certain cases of long-term experimentation this approach seems more reasonable than the minimax method which decides on the course of action appropriate to θ_n on the basis of x_n only,

and ignores the information about the prior distribution of ... in $\tau_1, \cdots, \tau_{n-1}$.

36. Smallest Average Confidence Sets for the Simultaneous Estimation of k Normal Means. (By Title). RAGHU RAJ BAHADUR, University of North Carolina.

Let $v = (x_{11}, \cdots, x_{1n_1}; \cdots; x_{k1}, \cdots, x_{kn_k})$ denote the combined samples of sizes $n_1, n_2, \cdots, n_k$ from normal populations $\pi_1, \pi_2, \cdots, \pi_k$, π_i having mean μ_i and variance σ_i^2. Writing $\mu = (\mu_1, \mu_2, \cdots, \mu_k)$, denote the k-dimensional Euclidean space of all points μ by R. Given any parameter point μ, $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_k)$, and any set-valued function $f(v)$ defined for all sample points v and having subsets of R as its values (which satisfies certain measurability hypotheses), let $\alpha(f \mid \mu, \sigma)$ = probability of the statement "$\mu \epsilon f(v)$" being false, and $\beta(f \mid \mu, \sigma)$ = expected Lebesgue measure of $f(v)$. We consider the problem of constructing $f(v)$ so as to make both α and β "as small as possible." One of the results obtained is as follows. Given p, $0 < p < 1$, let $f^0_{\lambda\,\zeta(p)}(v) = \{\mu : \Sigma_1^k n_i[(\bar{x}_i - \mu_i)/l_i]^2 < \zeta(p)\, \Sigma_1^k n_i[s_i/l_i]^2\}$, where $\bar{x}_i = n_i^{-1}\Sigma_1^{n_i} x_{ij}$, $s_i^2 = n_i^{-1}\Sigma_1^{n_i} (x_{ij} - \bar{x}_i)^2$, $\lambda = (l_1, l_2, \cdots, l_k)$, the l_i's being given positive constants, and $\zeta(p)$ being determined by $P(\chi_k^2 > \zeta(p)\, \chi_{N-k}^2) = p$, where χ_k^2, χ_{N-k}^2 are independent chi-square variables with k, $N - k$ degrees of freedom ($N = \Sigma_1^k n_i$). Then (a) obviously $\alpha(f^0_{\lambda\,\zeta(p)} \mid \mu, c\lambda) = p$ for all μ and all c, $0 < c < \infty$, and (b) if $f(v)$ is any other function such that $\alpha(f \mid \mu, c\lambda) \leq p$ for all μ and all c, either (i) $f(v)$ and $f^0_{\lambda,\zeta(p)}(v)$ differ by a set of measure zero for almost every v, or (ii) $\sup_{\mu \epsilon R} \{\beta(f \mid \mu, c\lambda)\} > \sup_{\mu \epsilon R} \{\beta(f^0_{\lambda,\zeta(p)} \mid \mu, c\lambda)\}$ for every c.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of general interest

Personal Items

Mr. Harry H. Goode, formerly head of the Special Projects Branch, Special Device Center, Office of Naval Research 1, New York, is now Supervisor of the Aero-Physics Group, Aeronautical Research Center, University of Michigan, Ann Arbor, Michigan.

Mr. William G. Howard, who was previously employed by the Johns Hopkins University, Institute for Cooperative Research, is presently employed as Mathematical Statistician in the Air Studies Division of the Library of Congress.

Miss Margaret Kampschaefer has accepted a position as Statistician in the U. S. Bureau of Labor Statistics, Minnesota Payroll Project, Minnesota Division of Employment and Security. She was formerly employed as Junior Mathematician at the Argonne National Laboratory, Naval Reactor Division, Chicago, Illinois.

Dr. Albert Noack has recently been appointed Professor of Actuarial Mathematics at the University of Koeln, Germany.

Second Berkeley Symposium on Mathematical Statistics and Probability

The Second Berkeley Symposium will be held at the Statistical Laboratory, University of California, Berkeley, from July 31 to August 12, 1950, with the

cooperation of the American Statistical Association (Biometrics Section), the Biometric Society (Western North American Region), the Econometric Society, the Institute of Mathematical Statistics, the Institute of Transportation and Traffic Engineering (UC), and the Office of Naval Research.

The Symposium will include sessions on mathematical statistics, probability, biometrics, econometrics, traffic engineering, astronomy, and physics. The complete program may be obtained from the Statistical Laboratory. The papers will be published by the University of California Press as the *Proceedings of The Second Symposium.*

Cumulative Index of Volumes 1–20

Attention is called to the fact that there is now available a cumulative index for Volumes 1 through 20 (1930–1949) of the *Annals of Mathematical Statistics.* Copies may be secured from the office of the Secretary-Treasurer for $1 00 per copy

New Members

The following persons have been elected to membership in the Institute

(December 1, 1949 to February 28, 1950)

Bain, John C., B A. (Univ of Toronto), President's Statistician, Abitibi Power & Paper Company, Ltd., 408 University Avenue, Toronto 2, Ontario, Canada.

Blakemore, George J., Jr., A.B. (George Wash Univ.), Student at George Washington University, *1748 Hobart St., N W , Washington 10, D C.*

Bross, Irwin, D. J., Ph.D. (North Carolina State College), Research Associate, Department of Biostatistics, School of Public Health, The Johns Hopkins University, 615 North Wolfe Street, Baltimore 5, Maryland.

Cansado Maceda, Enrique, Ph.D. (University of Madrid), Assistant Professor of Mathematical Statistics, Faculty of Sciences, University of Madrid and Official of the National Institute of Statistics, *Paseo de Rosales, 50 Madrid, Spain.*

Clatworthy, Willard H., M A (Univ of Kentucky), Student at the University of North Carolina, *Box 168, Chapel Hill, North Carolina*

Dinsmore, Robert J., A B (Univ. of Calif.), Student at the University of California, Berkeley, California, *2428 Milvia St., Berkeley 4, California.*

Enell, John W., Eng. Sc.D (New York Univ.), Assistant Professor of Administrative Engineering, New York University, *71 Ayers Court, West Englewood, New Jersey.*

Flores, Anna M., M.Sc (Univ. of Mexico), Mathematician, *Torres Adalid #511, Mexico City.*

Garner, Norman R., B.A (Univ of Rochester), Graduate Student at University of North Carolina, *15 Goldston Ave , Carrboro, North Carolina*

Hannan, James F., M.A (Harvard), Research Assistant, Department of Mathematical Statistics, University of North Carolina, P O Box 168, Chapel Hill, North Carolina.

Klein, Joseph, B.S. (Rutgers), Graduate Student at Rutgers University, *P O Box 501, Red Bank, New Jersey.*

Lewis, Evan J., Ph.D. (Cornell Univ), Physicist, Corning Glass Works, Corning, New York

Palekar, Madhukar N., B.S. (Bombay), Graduate Student in Department of Mathematical Statistics and Departmental Assistant, 108 Furnald Hall, Columbia University, New York 27, New York.

Page, Woodrow W., M.A. (Oklahoma Univ.), Graduate Student, University of North Carolina, *241 Jackson Circle, Chapel Hill, North Carolina.*

Pretorius, S. J., Ph.D (Univ. of London), Professor of Statistics, University of Stellenbosch, Soeteweide, Stellenbosch, Union of South Africa

Price, Don C., M.A. (Kent State Univ), Student, Department of Mathematical Statistics, University of North Carolina, *1621 Shorb Ave , N W , Canton 3, Ohio*

Scalora, Frank S., A.B. (Harvard), Assistant in Mathematics, 106 Mathematics Building, University of Illinois, Urbana, Illinois

Somerville, Paul N., B Sc. (Alberta, Canada), Graduate Student in Department of Mathematical Statistics, University of North Carolina, *316-B Dormitory, Chapel Hill, North Carolina.*

Sirken, Monroe G., M A. (Univ of Calif at L. A), Research Associate, Laboratory of Statistical Research, Department of Mathematics, University of Washington, Seattle, Washington

Stearn, Joseph L., M S. (College of N. Y), Mathematician, U S. Coast & Geodetic Survey, Department of Commerce, Washington, D. C.

Whelan, Walter J., M A. (Boston Univ), Student, Department of Mathematical Statistics, Columbia University, New York, *119 Wilmington Ave., Dorchester 24, Massachusetts*

Wile, Janet L., A.B (Univ of Rochester), Statistician, Department of Defense, Army and Transportation Corps, #156, 1813 Queens Lane, Arlington, Virginia.

Wilhelmsen, Lars, Aktuarkandidat (Oslo Univ.), Actuary, Storebrand, Boks 425, Oslo, Norway.

REPORT OF THE CHAPEL HILL MEETING OF THE INSTITUTE

The forty-second meeting of the Institute of Mathematical Statistics was held jointly with the Biometric Society (Eastern North American Region) at the Chapel Hill campus of the University of North Carolina on Friday, March 17, and Saturday, March 18, 1950 One hundred twenty-one persons registered, including the following members of the Institute·

R L Anderson, T W Anderson, Geoffrey Beall, C A Bennett, Mrs C A. Bennett, Nils Blomqvist, R C Bose, R A Bradley, Irwin Bross, Glen Burrows, L D Calvin, Uttam Chand, W. G Cochran, A C Cohen, Jr , W S Connor, Jr , P. P Crump, E E Cureton, R C Davis, W L Deemer, T G Donnelly, Churchill Eisenhart, J W. Fertig, S G. Ghurye, Leon Gilford, B. G Greenberg, F E Grubbs, Max Halperin, J F Hannan, Boyd Harshbarger, C R Henderson, Wassily Hoeffding, Harold Hotelling, A S. Householder,

W G. Howard, S L Isaacson, A. W. Kimball, Jr , B F Kimball, Marguerite Lehr, Guido Liserre, Eugene Lukacs, C L. Marks, H. A Meyer, Paul Minton, D J Morrow, Jack Moshman, C M. Motley, M. L. Norden, H. W Norton, Ingram Olkin, Paul Peach, J A Rafferty, Wyman Richardson, Jr , Herbert Robbins, S N Roy, S A Schmitt, R E. Serfling, D. H Shepard, P N Sommerville, E. W Stacy, J W Tukey, D F Votaw, Jr , F. M. Wadley, M. A. Woodbury, Marvin Zelen.

Professor R. L. Anderson presided at the opening session for contributed papers on Friday morning. The following papers were presented

1 *A Method of Estimating the Parameters of an Autoregressive Time Series.* Mr S G Ghurye, University of North Carolina.
2 *Most Powerful Rank Order Tests* Professor Wassily Hoeffding, University of North Carolina.
3. *The Comparison of Percentages in Matched Samples.* Professor W G. Cochran, Johns Hopkins University
4. *A Method of Estimating Components of Variance in Disproportionate Numbers* Professor H L Lucas, North Carolina State College
5 *On the Theory of Unbiased Tests of Simple Statistical Hypotheses Specifying the Values of Two Parameters.* Mr S L. Isaacson, Columbia University
6 *A Note on Orthogonal Arrays* Professor R C Bose, University of North Carolina.
7 *Transformations Related to the Angular and the Square Root* Mr. M F. Freeman and Professor J W Tukey, Princeton University
8 *Standard Inverse Matrices for Fitting Polynomials.* Mr F J. Verlinden, North Carolina State College

On Friday afternoon Dr. James A. Rafferty, School of Aviation Medicine, Randolph Field, Texas, gave an invited address on *Mathematical Models in Biology.* Professor Gertrude M. Cox then presided at a session for contributed papers, at which the following papers were presented:

1 *Small Sample Performance of Biological Statistics.* Mr Irwin Bross, Johns Hopkins University
2 *Methodology in the Study of Physical Measurements of School Children.* Professor B G Greenberg and Professor A. H. Bryan, University of North Carolina
3 *Tetrad Analysis in Yeast* Dr A S Householder, Oak Ridge National Laboratory
4 *Contribution to the Probabilistic Theory of Neural Nets I Randomization of Refractory Periods and of Stimulus Intervals.* Professor Anatol Rapoport, University of Chicago
5. *Theoretical and Experimental Aspects in the Removal of Airborne Matter by the Human Respiratory Tract* Professor H D. Landahl, University of Chicago. (Read by Professor Rapoport.)
6 *An Application of Biometrics to Zoological Classifications* Dr F M Wadley, Navy Department, Washington, D. C
7 *The Analysis of Hemotological Effects of Chronic Low-level Radiation* Mr. Jack Moshman, United States Atomic Energy Commission, Oak Ridge, Tennessee

A joint dinner of the two sponsoring organizations was held at the Carolina Inn on Friday evening, with an attendance of sixty-two Professor W G. Cochran as toastmaster introduced Chancellor R B House of the University of North Carolina who welcomed the gathering with words and music. Professor Gertrude M. Cox responded for the Biometric Society and Professor D F Votaw for the Institute

Professor Harold Hotelling presided at a Saturday morning symposium on

multivariate analysis Professor E. E. Cureton of the University of Tennessee gave the opening address on *Statistical Problems in Psychological Testing*. After a lively discussion the following contributed papers were presented:

1. *Accuracy of a Linear Prediction Equation in a New Sample* Professor George E Nicholson, Jr, University of North Carolina
2. *Independence of Quadratic Forms in Normally Correlated Variables* Professor Yukiyosi Kawada, Tokyo University of Literature and Science, Tokyo, Japan (Read by the chairman)
3. *Bounds on the Distribution of Chi-square.* Mr. S. A. Vora, University of North Carolina.

This was followed by a Biometric Society address by Professor C R. Henderson of Cornell University on *Estimation of Genetic Parameters.*

Professor W. G. Cochran presided at the final session for contributed papers on Saturday afternoon The following papers were presented:

1. *Estimating the Mean and Standard Deviation of Normal Populations from Double Truncated Samples* Professor A. C Cohen, Jr, University of Georgia
2. *Minimax Estimates of Location and Scale Parameters.* Mr. Gopinath Kallianpur, University of North Carolina.
3. *On Some Features of the Neyman-Pearson and Wald Theories of Statistical Inference, Their Interrelations and Bearing on Some Usual Problems of Statistical Inference.* Professor S. N Roy, University of North Carolina
4. *Note on Uniformly Best Unbiased Estimates* Mr. R C Davis, Naval Ordnance Test Station, Inyokern, Calif
5. *Competitive Estimation* Professor Herbert Robbins, University of North Carolina
6. *The Effect of an Unknown 'Location Disturbance' on "Student's" t Based on a Linear Regression Model* Professor Uttam Chand, Boston University
7. *Corrections for Non-normality for the Two-sample t and F distributions Valid for High Significance Levels* Professor Ralph A Bradley, McGill University
8. *Some Tests Based on the Empirical Distribution Function* Mr J F Hannan, University of North Carolina.
9. *On a Generalization of the Behrens-Fisher Problem.* (By title). Dr John E Walsh, Rand Corporation, Santa Monica, Calif
10. *Construction of Partially Balanced Designs with Two Accuracies.* (By title) Mr. S. S Shrikhande, University of North Carolina and Nagpur College, Nagpur, India
11. *Designs for Two-way Elimination of Heterogeneity.* (By title). Mr. S. S. Shrikhande
12. *Designs for Animal Feeding Experiments* (By title) Mr S. S Shrikhande
13. *A Truncated Sequential Procedure for Interval Estimation, with Applications to the Poisson and Negative Binomial Distributions* (By title) Mr D. Martin Sandelius, University of Washington and Uppsala University, Uppsala, Sweden
14. *A Generalization of the Method of Maximum Likelihood Estimating a Mixing Distribution* (By title) Professor Herbert Robbins, University of North Carolina
15. *Smallest Average Confidence Sets for the Simultaneous Estimation of k Normal Means.* (By title) Mr Raghu Raj Bahadur, University of North Carolina

About eighty-five members of the two organizations attended a tea given by Professor and Mrs. Hotelling at the conclusion of the Saturday afternoon session.

HERBERT ROBBINS
Assistant Secretary

FUNDAMENTAL LIMIT THEOREMS OF PROBABILITY THEORY[1]

By M. Loève[2]

University of California, Berkeley

no sooner is Proteus caught
than he changes his shape

1. Introduction. The fundamental limit theorems of Probability theory may be classified into two groups. One group deals with the problem of *limit laws* of sequences of sums of random variables, the other deals with the problem of *limits of random variables,* in the sense of almost sure convergence, of such sequences. These problems will be labelled, respectively, the Central Limit Problem (CLP) and the Strong Central Limit Problem (SCLP). Like all mathematical problems, the CLP and SCLP are not static, as answers to old queries are discovered they experience the usual development and new problems arise The development consists in (i) simplifying proofs and forging general tools out of the special ones (ii) sharpening and strengthening results (iii) finding general notions behind the results obtained and extending their domains of validity. *Analysis of this growth will put in relief the role and the interconnections of the fundamental limit theorems.*

Summary. The growth of the CLP for independent summands can be divided into three (overlapping) periods The first covers the Bernoulli case and the corresponding limit theorems of Bernoulli, de Moivre and Poisson. The first two theorems gave rise to the notions—from which the classical CLP stems—of the Law of Large Numbers (LLN) and of Normal Convergence (NC). Poisson's approach belongs to the set-up of the modern CLP.

The second period extends over two centuries and is devoted to the extension of the domains of validity of LLN and NC. This is the classical CLP period. Lyapunov's crucial work, submitted to the above treatment, led to the discovery of the natural boundaries of these domains by Lindeberg, Kolmogorov, Feller and P. Lévy.

However, the LLN and NC problems are but two particular cases of the general problem of limit laws of sequences of sums of independent random variables. The coming into sight and the solution of this problem—the third period of the CLP—covers less than ten years. The tools forged for the classical CLP proved to be powerful enough and the final solution is due to P. Lévy, Khintchine, Gnedenko and Doeblin.

[1] This paper was presented to the New York meeting of the Institute of Mathematical Statistics on December 27, 1949.

Editor's Note The Institute of Mathematical Statistics has formed a Committee on Special Invited Papers to invite lecturers to deliver expository addresses to the Institute with the understanding that the Special Invited Papers are to be published in the *Annals of Mathematical Statistics* This paper is the first one invited by the Committee.

[2] This work is supported in part by the Office of Naval Research

The CLP for dependent variables started with so called Markoff chains. The study of their limit properties is due essentially to Markov, S Bernstein and Doeblin. For more general forms of dependence the LLN and NC problems were investigated by P. Lévy and Loève after the crucial work of S. Bernstein The modern CLP was considered only recently (Loève).

The SCLP stems from the strengthening by Borel of the Bernoulli theorem and the sharpening of Borel's result by Khintchine. They gave rise to the notions of Strong Law of Large Numbers (SLLN) and of the Law of the Iterated Logarithm (LIT).[3] The domains of validity were extended to their boundaries by Kolmogorov, P Lévy and Feller. In the case of dependence, results are due to G. D Birkhoff, P. Lévy, W Doeblin, and Loève However, the SCLP has not attained, at present, the harmonious development of the CLP.

Notations. Let $\mathfrak{L}(X)$ be the *law* of a (real) random variable (r v.) X The law is defined by the *distribution function* (d.f) $F(x) = P(X < x)$. As is well known $\mathfrak{L}(X)$ is determined by the *characteristic function* (ch. f.)

$$f(u) = \int_{-\infty}^{+\infty} e^{iux}\, dF(x), \qquad -\infty < u < +\infty.$$

When a r.v possesses subscripts, the same subscripts will be used for its d.f. and ch f EX will denote the *expectation* of X:

$$EX = \int_{-\infty}^{-\infty} x\, dF(x),$$

and $\sigma^2(X)$ will denote the *variance* of X:

$$\sigma^2(X) = E(X - EX)^2.$$

With a random event A we associate a r.v., to be called *indicator* of the event A, which takes values 1 and 0 respectively, according as A occurs or does not occur. If X is the indicator of an event A of probability p, then $EX = p$ and $\sigma^2(X) = pq$, where $q = 1 - p$. To avoid trivialities we shall assume that $pq \neq 0$.

Two laws $\mathfrak{L}(X_1)$ and $\mathfrak{L}(X_2)$ will be said to belong to the same *complete type* if there exist two numbers $a \neq 0$ and b such that $P\{X_1 \leqq x\} = P\{aX_2 + b \leqq x\}$ If values of a are restricted to positive values, then the two laws are said to belong to the same *type*. If two independent r.v.'s obey $\mathfrak{L}$ and their sum belongs to the type of $\mathfrak{L}$, then $\mathfrak{L}$ and its type are said to be *stable*. Three classes of laws play an essential role in the CLP: the normal and the degenerate types and the Poisson complete types.

$\mathfrak{N}(m, \sigma)$ is a *normal* law if it is defined by

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(1/2\sigma^2)(t-m)^2}\, dt \qquad (\sigma > 0).$$

[3] For a very thorough and deep analysis of the NC and LIT problems and their solutions see FELLER, *Bull. Am. Math. Soc*, Vol. 51 (1945), pp. 800-832, under the same title as that of the present paper.

$\mathfrak{L}(m)$ is a law *degenerate* at m, if it attaches probability 1 to the value m. $\mathcal{P}(\lambda; a, b)$ is a *Poisson* law if

$$P(X = ak + b) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad (\lambda > 0), \qquad k = 0, 1, 2, \cdots;$$

the familiar Poisson law is $\mathcal{P}(\lambda; 1, 0)$.

A law $\mathfrak{L}(X_n)$ is said to converge to the law $\mathfrak{L}(X)$ as $n \to \infty$, if $F_n(x)$ converges to $F(x)$ at the continuity points of the latter. In this paper, all limits will be considered for $n \to \infty$, if not otherwise stated.

The structure of sequences of r.v.'s whose limit properties are investigated will be called the *limiting process* of the problem. The limiting process of *sequences of sums* is that of sequences of the form $S_{n,\nu_n} = \sum_{k=1}^{\nu_n} X_{n,k}$, where $\nu_n \to \infty$. The limiting process of *normed sums* is that of sequences of the form $\frac{S_n}{a_n} - b_n$ with $S_n = \sum_{k=1}^{n} X_k$, where $a_n > 0$ and b_n are real numbers. Normed sums are a special form of sequences of sums: take $\nu_n = n$, $X_{n,k} = \frac{X_k}{a_n} - \frac{b_n}{n}$, then $S_{n,\nu_n} = \frac{S_n}{a_n} - b_n$.

To avoid repetitions we shall note, once and for all, that limit types rather than limit laws appear in the case of normed sums, because, if $\mathfrak{L}(X)$ is their limit law, then any law of its type is obtainable as a limit law by a convenient change of origin b_n and of scale a_n, independent of n. The importance of the notion of type is due, primarily, to this property. In fact, even more is true: *if $\mathfrak{L}(X_n)$ converges to $\mathfrak{L}(X)$ and $\mathfrak{L}(a_nX_n + b_n)$ converges to $\mathfrak{L}(Y)$, then $\mathfrak{L}(X)$ and $\mathfrak{L}(Y)$ belong to the same type, provided neither is degenerate* (Khintchine [20])

I. Central Limit Problem

2. Origin of the CLP: Binomial case. Three limit theorems are at the origin of the CLP; the first, due to Bernoulli ([2], 1713), laid the ground. *Let S_n be the number of occurrences of an event A of probability p in n identical and independent trials. Then, for every $\epsilon > 0$,*

$$P\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \to 0.$$

Bernoulli found this result by a direct, but cumbersome, analysis of the behaviour of the binomial probabilities

$$P\{S_n = k\} = C_n^k p^k q^{n-k}, \qquad k = 0, 1, \cdots, n.$$

Sharpening this analysis, de Moivre ([7], 1730) obtained the second limit theorem of probability theory which, in the form given to it by Laplace, states that: *For every x*

$$P\left\{\frac{S_n - np}{\sqrt{npq}} < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt.$$

Suppose now, with Poisson ([36], 1837), that the probability $p = p_n$ depends upon the number n of trials and, more precisely, that $p_n = \frac{\lambda}{n}$, where λ is a positive constant. Write then $S_{n,n}$, instead of S_n, for the number of occurrences of the considered event in a group of n trials. By a direct analysis of the binomial probabilities, much easier to carry out than the preceding ones, it follows that for $k = 0, 1, \cdots$,

$$P\{S_{n,n} = k\} \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

Let X_k be the indicator of the event A in the k-th trial. The number of occurrences S_n is the sum $\sum_{k=1}^{n} X_k$ of n of these independent and identically distributed indicators. The first two limit theorems mean that

$$\mathfrak{L}\left(\frac{S_n - ES_n}{n}\right) \to \mathfrak{L}(0) \quad \text{and} \quad \mathfrak{L}\left(\frac{S_n - ES_n}{\sigma S_n}\right) \to \mathfrak{N}(0, 1).$$

Thus we have two limiting processes, (both special and completely specified forms of normed sums), and two limit laws (more precisely two limit types, see introduction), a degenerate and a normal one

Poisson's limiting process is utterly different. $S_{n,n}$ is still a sum $\sum_{k=1}^{n} X_{n,k}$ of independent and identically distributed indicators but, as n varies, *all* $X_{n,k}$ change, $P(X_{n,k} = 1) = \frac{\lambda}{n}$ and

$$\mathfrak{L}(S_{n,n}) \to \mathcal{P}(\lambda, 1, 0).$$

While the two first theorems with their special limiting processes and limit laws played a central role in the development of Probability theory, Poisson's result stood isolated and ignored until about fifteen years ago[4] We shall see further that there was a deep reason for its isolation and also that, surprisingly enough, Poisson laws are, in a sense, more fundamental for the CLP, than the normal law

3. The classical CLP and its extension. From the time of Laplace until 1935, research in the domain of limit laws was centered about the extension to summands other than indicators of the validity of the two first limit theorems This is the period of the classical CLP: *Let* $S_n = \sum_{k=1}^{n} X_k$ *be sums of independent r.v.'s. Find necessary and sufficient conditions for the LLN and for NC*, i e., conditions under which, respectively,

$$\text{LLN: } \mathfrak{L}\left(\frac{S_n - ES_n}{n}\right) \to \mathfrak{L}(0),$$

$$\text{NC: } \mathfrak{L}\left(\frac{S_n - ES_n}{\sigma(S_n)}\right) \to \mathfrak{N}(0, 1).$$

[4] In Uspensky's textbook (1937!) Poisson's law is mentioned once—in an exercise.

It is assumed that EX_k's and EX_k^2's exist. The d.f. not being completely specified as in the Bernoulli case, the direct Bernoulli-de Moivre approach is of no avail and general methods are necessary. The first to appear was the method of moments relative to bounds of d.f. in terms of their moments (Tchebicheff [40], Markov [37]). The relation

$$P\left\{\frac{|S_n - ES_n|}{n} > \epsilon\right\} \leqq \frac{\sigma^2(S_n)}{\epsilon^2 n^2}, \qquad \epsilon > 0,$$

together with

$$\sigma^2(S_n) = \sum_{k=1}^{n} \sigma^2(X_k),$$

entails at once a LLN theorem (Tchebicheff-Markov): *If*

$$\frac{1}{n^2} \sum_{k=1}^{n} \sigma^2(X_k) \to 0,$$

then the LLN holds.

This result can be easily improved (bringing it into closer analogy with Lyapunov's theorem): *If there exists a constant* $\delta > 0$ *such that*

$$\frac{1}{n^{1+\delta}} \sum_{k=1}^{n} E\,|X_k - EX_k|^{1+\delta} \to 0$$

then the LLN holds.

It contains then a Markov's LLN condition: LLN holds if $E\,|X_k - EX_k|^{1+\delta} \leqq C$ where C is independent of k.

In a much more elaborate form the method of moments gives also a NC theorem (Tchebicheff-Markov): *If* $EY_n^k \to EZ^k$ *for* $k = 1, 2, \cdots$, *and* $\mathfrak{L}(Z) = \mathfrak{N}(0, 1)$, *then* $\mathfrak{L}(Y_n) \to \mathfrak{N}(0, 1)$

This theorem has been extended to more general limit laws. However the inherent defects of the method of moments remain. Even if moments of all orders exist, they do not necessarily determine a unique d.f. A definitive result in this direction is the Fréchet-Shohat theorem: *If* $EY_n^k \to m^{(k)}$ *for all* k, *there exists a subsequence* $\mathfrak{L}(Y_{n_\nu})$ *which converges to a limit law* $\mathfrak{L}$ *with moments* $m^{(k)}$. *Moreover, if the moment problem is determined, i.e., if the* $m^{(k)}$ *determine a unique law, then the whole sequence* $\mathfrak{L}(Y_n)$ *converges to* $\mathfrak{L}$.

To apply the convergence theorem to the NC part of the classical CLP, one has to assume existence of moments of all orders. In particular, it does not seem suitable for proving Lyapunov's theorem. Yet, the simple *truncation* idea (Markov) not only overcomes this seemingly insurmountable obstacle, but also provides a method per se. It associates with the summands X_k "truncated" r.v.'s X_k'; for $k \leqq n$ and c_n conveniently chosen real numbers,

$$X_k' = X_k \text{ if } |X_k| \leqq c_n,$$

$$X_k' = 0 \quad \text{if } |X_k| > c_n.$$

Nevertheless, the method of moments is too cumbersome and was soon to be discarded in favor of that of ch.f.'s.

The turning point for the entire CLP is Lyapunov's introduction of the *method of ch.f.'s*. The ch.f.'s were well known and used already by Laplace. However, the first convergence property, proved but not stated, is due to Lyapunov [28]: *If the ch.f.'s $g_n(u)$ of $\mathfrak{L}(Y_n)$ converge to the ch.f. $e^{-u^2/2}$ of $\mathfrak{N}(0, 1)$, then $\mathfrak{L}(Y_n) \to \mathfrak{N}(0, 1)$.* From it he deduced the first general NC theorem [28, 29]: *If there exists a number $\delta > 0$ such that*

$$\frac{1}{\sigma^{2+\delta}(S_n)} \sum_{k=1}^{n} E \mid X_k - EX_k \mid^{2+\delta} \to 0,$$

then NC holds.

The ch.f. became, in the hands of P. Lévy [21], a general tool, instrumental in the subsequent tremendous growth of the CLP, with the so called

CONTINUITY THEOREM. *If the ch.f.'s $g_n(u)$ converge to a function $g(u)$ continuous at $u = 0$, then $\mathfrak{L}(Y_n)$ converge to a limit law $\mathfrak{L}$ and $g(u)$ is its ch.f.; and conversely.*

The methods of ch.f. and of truncation dominate at present the limit problems of Probability theory.

In spite of the generality of the above conditions for LLN and NC, they are not *necessary* conditions. In fact they are not sharp enough since they assume the existence of moments of higher order than those which figure in the classical CLP. However the tools forged proved powerful enough to get its complete solution. The truncation method yielded to Kolmogorov ([16, 1928) the complete answer to the LLN problem. A "smoothing" device, due to Lyapunov, provided Lindeberg ([20], 1922) with adequately sharp sufficient conditions, using ch.f.'s P. Lévy ([22], 1922) proved Lindeberg's result and Feller ([11], 1935) showed that, under a natural restriction, these conditions are also necessary.

Solution of the classical CLP.

1. *LLN holds if, and only if,*

$$\sum_{k=1}^{n} \int_{|x|>n} dF_k(x + EX_k) \to 0 \quad \textit{and} \quad \sum_{k=1}^{n} \frac{1}{n^r} \int_{|x|<n} x^r \, dF_k(x + EX_k) \to 0 \quad \textit{for } r = 1, 2.$$

2. *NC holds and* $\max_{k \leq n} \dfrac{\sigma(X_k)}{\sigma(S_n)} \to 0$ *if, and only if, for every $\epsilon > 0$,*

$$\sum_{k=1}^{n} \frac{1}{\sigma^2(S_n)} \int_{|x|>\epsilon\sigma(S_n)} x^2 \, dF_k(x + EX_k) \to 0.$$

An unsatisfactory feature of the classical CLP is the assumption, made at the start, of existence of certain moments. They are used to avoid, as $n \to \infty$, the shift, towards infinite values, of the probability spread by changing the origin and the scale of values of S_n. However there is no specific reason for these special choices of norming quantities a_n and b_n except that, historically,

they appeared as a straightforward extension of Bernoulli and de Moivre ones. Moreover, even if these moments do not exist, there is no reason not to try to find norming quantities. (Take X_k's to be independent and identically distributed as follows: to $\pm\sqrt{m}$ where $m = 1, 2, \cdots$, attach probabilities $\frac{3}{\pi^2 m^2}$. The second moments are infinite, yet norming S_n by $c\sqrt{n \log n}$, we have NC.) Thus the CLP becomes the problem of the LLN and NC for general normed sums $\frac{S_n}{a_n} - b_n$.

The extended classical NC problem was solved, masterfully and independently, by Feller ([10], 1935) using ch f.'s and by P. Levy ([25], 1935) who applied the method of truncation The extension of the results to the more general set-up of the following section is trivial and will be given there. Feller also solved ([11], 1937) the extended LLN problem.

In this new set-up a question arises at once. *Given the r v 's X_k, do there exist numbers which will produce the desired convergence? If so, how can they be found?* This problem is perhaps more difficult than the previous one and is specifically linked with the limiting process of normed sums We shall give here a criterion, due to Feller ([10], 1935), which solves entirely the NC problem.[5] Take as origin of values of the summands their medians and let $c_n(\epsilon)$ be the g.l b. of the x's for which $\sum_{k=1}^n P(|X_k| > x) \leq \epsilon$. Then *norming quantities a_n and b_n such that $\mathfrak{L}\left(\frac{S_n}{a_n} - b_n\right) \to \mathfrak{N}(0, 1)$ and $\max_{k \leq n} P\left\{\left|\frac{X_k}{a_n} - \frac{b_n}{n}\right| > \epsilon\right\} \to 0$ exist if, and only if, for every $\epsilon > 0$,*

$$\frac{1}{c_n^2(\epsilon)} \int_{|x| < c_n(\epsilon)} x^2 \, dF_k(x) \to \infty.$$

4. Modern CLP. At the same time that the classical CLP neared its happy end, a new and much wider problem of limit laws appeared and, because the necessary tools were at hand, was solved almost at once. Various particular problems, of which the classical CLP is one, contributed to its set-up

Since the discovery, in the Bernoulli case, of the LLN and NC, the problem of limit laws has been centered about extensions of their domains of validity for more and more general normed sums. A similar query about the Poisson convergence would have provided us with a new problem. As soon as we drop the restriction that in $S_{n,\nu_n} = \sum_{k=1}^n X_{n,k}$ the r.v.'s $X_{n,k}$ are indicators, we are led to the problem of finding conditions under which laws of sums of independent r.v.'s will converge to a Poisson law. We have here not only a different limit law than in the CLP but also a more general limiting process. An utterly different problem, stated and solved by P. Lévy [21], is the following: *find the*

[5] As for the LLN, norming numbers, such that the LLN holds always exist whatever be the r v 's X_k Hence, from the point of view of limit types of normed sums, the degenerate type is to be considered as a degenerate form of every limit type.

possible limit laws of normed sums of independent and identically distributed r v.'s (the answer is that *they are the stable laws*). For the first time one does not inquire about a completely specified limit law but about the class of *all* limit laws for a fairly general limiting process. Thus, starting with limit theorems with completely specified limiting processes and limit laws, after two centuries of struggle Probability theory got rid of initial restrictions.

The general set-up is now visible. The limiting process is that of sequences of sums of independent r.v.'s. The queries are about the classes of possible limit laws and conditions of convergence. However, so general a limit problem is without content. In fact, the limiting process is that of arbitrary sequences of r v.'s. let $\{Y_n\}$ be any sequence of r v.'s and take $X_{n,1} = Y_n$, $\mathfrak{L}(X_{n,k}) = \mathfrak{L}(0)$ for $k > 1$. Any law $\mathfrak{L}$ belongs to the class of limit laws: take $\mathfrak{L}(Y_n) \equiv \mathfrak{L}$ Hence some restriction is needed. To find a "natural" restriction consider the previous problems. Their common feature is that the limiting process is that of sequences of *sums* of independent r.v.'s, *the number of summands increasing indefinitely*. If we wish to emphasize this feature, a relatively small number of summands ought not to have a preponderant role in the determination of the limit laws A "natural" restriction is then a requirement of *uniform asymptotic negligibility* (uan) *of the summands*, i e., for every $\epsilon > 0$, $P\{\,|X_{nk}| > \epsilon\} \to 0$ uniformly in k. We come thus to the *Modern CLP. Let* $S_{n,\nu_n} = \sum_{k=1}^{\nu_n} X_{n,k}$, $\nu_n \to \infty$, *be sums of r.v 's* $X_{n,k}$, *mutually independent for every fixed n, and such that*

$$\max_k P\{X_{n,k}\,| > \epsilon\} \to 0;$$

characterize the class $\{D\}$ *of limit laws of the* S_{n,ν_n} *and find necessary and sufficient conditions for convergence to any element of this class.*

The solution of this problem is essentially due to the results of investigation of random functions $X(t)$ with independent increments. Let $X(0) = 0$, divide the interval $(0, t)$ into ν_n subintervals (t_{k-1}, t_k) with $t_0 = 0$, and denote by X_{nk} the increment $X(t_k) - X(t_{k-1})$. Then $X(t) = \sum_{k=1}^{\nu_n} X_{nk}$ where X_{nk} are independent r.v.'s. If, moreover, $X(t)$ is continuous in probability for every t, i.e., if $\mathfrak{L}\{X(t+h) - X(t)\} \to \mathfrak{L}(0)$ as $h \to 0$, then the $X_{n,k}$ can be chosen to obey the uan restriction as $\nu_n \to \infty$ Hence $\mathfrak{L}\{X(t)\}$ might be expected to belong to $\{D\}$.

The particular case of the modern CLP for summands and limit laws with the finite second moments was solved by Bawly [1], using Kolmogorov's characterization of $X(t)$'s with finite second moments [7]. The general problem, thanks to a much more general result by P. Lévy ([24], 1934), was solved by P. Lévy, Khintchine ([20], 1937), Gnedenko ([14], [15], 1938, 1939) and Doeblin ([8], 1938–1939). The method used throughout was that of ch.f.'s. (except in the case of Doblin who used also the P. Lévy "dispersion" function).

One can avoid an explicit introduction of the considered random function $X(t)$, limiting oneself to the corresponding (infinitely divisible) laws. For a very large n, S_{n,ν_n} is, roughly speaking, a very large number ν_n of very small (in probability) independent summands. This leads at once to the consideration

of laws which possess such a property for any ν_n and, first, the *infinitely divisible* (i.d.) laws. A law is i.d if it is a law of sums of an arbitrarily large number of independent and identically distributed r v.'s. In other words, *$f(u)$ is the ch.f of an i d. law if $[f(u)]^{1/n}$ is a ch f. for every positive integer n.* One might expect i.d. laws to belong to $\{D\}$ and, surprisingly enough, it turns out that, because of the uan, $\{D\}$ contains *only* i.d. laws.

We can now state the solution of the modern CLP, in three parts. Let $\fint_{-a}^{+a} = \int_{-a}^{-0} + \int_{+0}^{+a}$, let $\phi(x)$ be any function, defined and non-decreasing in $(-\infty, -0)$ and $(+0, +\infty)$, with $\phi(-\infty) = \phi(+\infty) = 0$ and $\fint_{-\epsilon}^{+\epsilon} x^2\, d\phi(x) < \infty$, and let α and β be real numbers.

I. *The function $f(u)$ is the ch.f. of an i.d. law if, and only if,*

$$\log f(u) = i\alpha u - \frac{\beta^2}{2} u^2 + \fint_{-\infty}^{+\infty} \left(e^{iux} - 1 - \frac{iux}{1 + x^2}\right) d\phi(x),$$

and $f(u)$ determines uniquely α, β and $\phi(x)$ at all the continuity points of the latter• (P. Lévy).

Normal laws are obtained for $\phi(x) \equiv 0$ and Poisson laws correspond to the $\phi(x)$ with one point of increase ($x \neq 0$) only. The fundamental role of Poisson laws appears clearly since, roughly speaking, an i d. law is the convolution of a normal law and a continuum of Poisson ones. This role is further emphasized by the following theorem (Khintchine [20]): *A law is i d. if, and only if, it is the limit law of sequences of sums of independent Poisson r.v.'s* In other words, the class of i.d laws is the closure of laws of finite sums of independent Poisson r.v.'s.

II. *The class $\{D\}$ of limit laws of the modern CLP coincides with that of i d. laws* (P Lévy-Khintchine).

Together with I this result characterizes in an explicit manner the class $\{D\}$. An immediate question arises (Khintchine). What about the limit laws of normed sums? The answer is the following (P. Lévy [27]). Let $y = \log |x|$, $\psi_1(y) = -\phi(x)$ for $x < 0$, $\psi_2(y) = \phi(x)$ for $x > 0$ where $y = \log |x|$. *The limit laws of normed sums, under uan, are the i d. laws with convex $\psi_k(y)$, $k = 1, 2$.*

In particular a Poisson law does not belong to this subclass $\{D_N\}$ of $\{D\}$, hence cannot be obtained as a limit law of normed sums. This brings out the deep reason for the isolation in which the Poisson law remained as long as the limiting process was restricted to that of normed sums [6] II shows that, with respect to the possible limit laws, the limiting process of the modern CLP is definitely wider than that of the classical CLP and of its extension. However the entire class $\{D\}$ can be obtained with normed sums, provided we consider

---

[6] A problem, specific for normed sums, arises· given r.v.'s X_k, find necessary and sufficient conditions for existence of norming numbers such that the laws of normed sums would converge to a given element of $\{D_N\}$ and, if they exist, find them. Feller's NC criterion solves a particular case of this problem.

not only limit laws but also "accumulation" laws (P. Lévy-Khintchine): *A law is i.d. if, and only if, it is the limit law of a subsequence of normed sums of independent and identically distributed r.v.'s.*

I and II provided Gnedenko and, independently, Doeblin with the properties which allowed them to find conditions of convergence, thus completing the solution of the modern CLP. Let

$$\sigma_\epsilon^2(X) = \int_{|x|<\epsilon} x^2\, dF(x) - \left[\int_{|x|<\epsilon} x\, dF(x)\right]^2$$

denote a "truncated" variance of X.

III. *Under uan, $\mathfrak{L}(S_{n,\nu_n} - b_n)$ converges, necessarily to an i.d. law "for a convenient choice of b_n", if, and only if,*

$$\text{(i)} \quad \sum_{k=1}^{\nu_n} F_{nk}(x) \to \phi(x) \text{ for } x < 0, \qquad \sum_{k=1}^{\nu_n} [1 - F_{nk}(x)] \to -\phi(x) \text{ for } x < 0$$

at the continuity points of $\phi(x)$, and

$$\text{(ii)} \qquad \lim_{\epsilon\to 0} \liminf_{n} \sum_{k=1}^{\nu_n} \sigma_\epsilon^2(X_{n,k}) = \beta^2 .$$

In particular, since normal laws correspond to $\phi(x) \equiv 0$, the NC conditions of Feller and P. Lévy follow: *$\mathfrak{L}(S_{n,\nu_n} - b_n)$ converges to $\mathfrak{N}(0, 1)$ for a convenient choice of b_n and uan holds if, and only if, for every $\epsilon > 0$,*

$$\text{(i)} \quad \sum_{k=1}^{\nu_n} \int_{|x|>\epsilon} dF_{nk}(x) \to 0 \qquad \text{and} \qquad \text{(ii)} \quad \sum_{k=1}^{\nu_n} \sigma_\epsilon^2(X_{nk}) \to 1.$$

The first condition shows that among all limit laws under uan, limit normality corresponds to a sufficiently strong asymptotic negligibility of the summands, and, more precisely, to

$$\sum_{k=1}^{\nu_n} P(|X_{nk}| > \epsilon) \to 0,$$

or, equivalently, to

$$P(\max_k |X_{nk}| > \epsilon) \to 0.$$

Another illuminating characterization of NC (Raikov [39]) follows also from III. Take for origin of values of summands the truncated first moments $\int_{|x|<1} x\, dF_{nk}(x)$. Then $\mathfrak{L}(S_{n,\nu_n} - b_n) \to \mathfrak{N}(0, 1)$ *for a convenient choice of b_n if, and only if,* $\mathfrak{L}(\sum_{k=1}^{\nu_n} X_{nk}^2) \to \mathfrak{L}(1)$.

5. CLP in the case of dependence. Limit problems for sums of dependent r.v.'s. were considered for the first time by Markov [37], less than fifty years ago. He extended the first two limit theorems of probability theory to the case of events linked in *chain*, i.e., such that $P(A_k \mid A_1, \cdots A_{k-1}) = P(A_k \mid A_{k-1})$.

However the crucial work in this field is the celebrated memoir by S. Bernstein ([3], 1927) which has the same historical importance for the dependence case as that of Lyapunov has for the classical CLP

Let $\{X_k\}$ be a sequence of r.v.'s. $E'X_k$ will denote the conditional expectation of X_k, given $X_1, \cdots X_{k-1}$. Consider the sequence of sums $S_n = \sum_{k=1}^{n} X_k$, with $EX_k \equiv 0$ and let $\sigma_n = \sqrt{\sum_{k=1}^{n} \sigma^2(X_n)}$

BERNSTEIN'S NC THEOREM. *If*

$$\text{(i)} \quad \frac{1}{\sigma_n}\sum_{k=1}^{n} sup \mid E'X_k \mid \to 0, \qquad \text{(ii)} \quad \frac{1}{\sigma_n^2}\sum_{k=1}^{n} sup \mid E'X_k^2 - EX_k^2 \mid \to 0,$$

and

$$\text{(iii)} \quad \frac{1}{\sigma_n^3}\sum_{k=1}^{n} sup\ E' \mid X_k \mid^3 \to 0,$$

then

$$\mathfrak{L}\left(\frac{S_n}{\sigma_n}\right) \to \mathfrak{N}(0, 1).$$

Obviously, if the X_k's are independent, this theorem reduces to Lyapunov's with $\delta = 1$. The method used is still that of ch.f 's From this result Bernstein deduces various particular NC cases and, applying them to Markov chains, extends the latter's results.

The unpleasant feature of the above theorem is the use of suprema of conditional expectations and, except when the r v.'s X_k are bounded, one cannot exfect these suprema to be finite. On the other hand, the conditional expectations are r v.'s and it would be natural to associate their values with the corresponding probabilities This can be done and Bernstein's theorem can be improved in various directions simultaneously. First it may be stated for sequences of sums S_{n,ν_n}—this is trivial; next it extends to $\delta > 0$ instead of $\delta = 1$—this contains completely Lyapunov's result but is of secondary interest. Then NC can be replaced by *asymptotic normality*, i.e., by the existence of a sequence of normal laws $\mathfrak{N}(0, \sigma_n)$ such that the "distance" between $\mathfrak{L}(S_{n,\nu_n})$ and $\mathfrak{N}(0, \sigma_n)$ would approach zero as $n \to \infty$—this is quite simple to get. However, significant improvements are obtained on replacing suprema by expectations. Let $F_n(x)$ be the d.f. of S_{n,ν_n} and $G_n(x)$ be that of $\mathfrak{N}(0, \sigma_n^2)$. Then, taking $EX_{nk} \equiv 0$, we have the following

NC THEOREM. *If* (i) $\sum_k E \mid E'X_{nk} \mid \to 0$, (ii) $\sum_k E \mid E'X_{nk}^2 - EX_{nk}^2 \mid \to 0$ *and* (iii) *there exists a constant* $\delta > 0$ *such that* $\sum_k \mid X_{nk} \mid^{2+\delta} \to 0$, *then* $F_n(x) - G_n(x) \to 0$.

This theorem shows that, so far as moments of order higher then the second are concerned, the NC condition is the same as in the case of independence In this last case the theorem is a slight improvement of that of Lyapunov. In 1941 condi-

tions for LLN and NC were given (Loève [31], [32]) in the frame of the modern CLP, without assuming the existence of moments; when independence is assumed, they reduce to those given by Feller. Conditions for NC which in the case of independence, reduce to Lindeberg's, were then deduced in the particular case of finite second moments and special cases of NC, including those considered by S Bernstein, were obtained.

The whole modern CLP had not been considered until lately (Loève, [33–35]). It appeared useful to extend the CLP to an "Asymptotic Central Problem" (ACP); primarily, to the behavior of $\mathfrak{L}(S_{n,\nu_n})$ as $n \to \infty$. This in turn, led to the introduction of laws "in a wide sense," i.e., with possible positive probabilities for infinite values. To the sequence $\{\mathfrak{L}(S_{n,\nu_n})\}$ is associated another conveniently chosen sequence $\mathfrak{L}_n$ of laws of sums; if $\mathfrak{L}_n \to \mathfrak{L}$ or $\mathfrak{L}_n \equiv \mathfrak{L}$ then the ACP reduce to the CLP. The investigation uses an extension of the P. Lévy convergence theorem for ch f.'s and the modern CLP solutions are obtained as particular cases The case of sums of a random number of r.v.'s,[7] as well as the multidimensional case, are easily treated by the same methods [35]

Many new problems arise in ACP. The foremost corresponds to possible relaxations of the uan condition For instance, in the case of independence, the relaxed condition

$$\max_k P\{| X_{nk} - Y_k | > \epsilon\} \to 0, \qquad \text{for every } \epsilon > 0,$$

where $Y_1, Y_2, \cdots$ are independent, does not change, essentially, the nature of the ACP Yet, as soon as dependence is introduced, the whole outlook changes and it would be interesting to investigate various new possibilities which thus arise. On the other hand, stricter than uan conditions are of special interest when independence is not assumed The one which seems natural is the following:

$$\max_k \sup P'\{| X_{nk} | > \epsilon\} \to 0, \qquad \text{for every } \epsilon > 0,$$

where $P'(A_{nk})$ denotes the conditional probability of the event $A_{n,k}$, given $X_{n,1}, \cdots, X_{n,k-1}$. An immediate problem is whether this or an analogous restriction enables us to find, not only sufficient, but also necessary conditions for various convergences and various cases of dependence.

II. The Strong Central Limit Problem

6. The Bernoulli case and its extension. A sequence $\{X_n\}$ such that the corresponding sequence of laws converges does not, in general, determine a r.v. X which might be considered, in some sense, as the limit of X_n. However, if we define two r.v 's X and X' such that $P(X \neq X') = 0$ as equivalent, then, whenever $\mathfrak{L}(X_m - X_n) \to \mathfrak{L}(0)$ as $\dfrac{1}{m} + \dfrac{1}{n} \to 0$, the sequence $\{X_n\}$ determines a

[7] H Robbins (*Bull. Am Math Soc*, Vol. 54 (1948), pp. 1151–1161. studied in detail the case of independent and identically distributed X_k's with $EX_k^2 < \infty$ and ν_n, independent of X_k's, with $E_n^2 < \infty$

unique r.v. X (up to an equivalence)—for which $P\{|X_n - X| > \epsilon\} \to 0$ for every $\epsilon > 0$. This X is the limit *in probability* of X_n.

Yet, an observed sequence of values of $\{X_n\}$ need not converge to the observed value of X. For instance, let Y be a r.v. uniformly distributed over $(0, 1)$. Consider the sequence $\{D_n\}$ of partitions of $(0, 1)$ into n equal subintervals and to the k-th subinterval of D_n attach the indicator $X_{n,k}$ of the event when Y falls within this subinterval. The sequence $X_{1,1}$; $X_{2,1}$, $X_{2,2}$; $X_{3,1}$, $X_{3,2}$, $X_{3,3}$, $\cdots$ converges in probability to zero since $P(X_{nk} \neq 0) = \frac{1}{n}$, for $k = 1, 2, \cdots, n$, approaches zero as $n \to \infty$. On the other hand, observed values of X_{nk}'s, for $k = 1, 2, \cdots, n$, will contain $n - 1$ zeros and a one, except in cases of total probability zero. Hence, except in these cases, any observed sequence will contain infinitely many zeros and infinitely many ones and will not converge.

The Bernoulli theorem means only that $f_n = \frac{S_n}{n}$ converges in probability to zero. Borel showed, in a fundamental memoir ([5], 1909), that Bernoulli's statement is too weak, and, in fact, that observed values of f_n converge to zero, except in cases of total probability zero. Borel's proof is based upon a direct analysis of the de Moivre-Laplace approach to NC. Thus a new domain in probability theory was opened to exploration.

FIRST STRONG LIMIT THEOREM. *In the Bernoulli case*

$$P\{\lim_n f_n = p\} = 1.$$

This leads to the introduction in probability theory of the notion of almost sure (a.s.) convergence:

$$X_n \xrightarrow{\text{a.s.}} X \text{ if } P\{\lim_n X_n = X\} = 1,$$

or, equivalently, if for every $\epsilon > 0$,

$$P\{\,|X_{n+k} - X| > \epsilon \text{ for } k = 1, \quad \text{or } 2 \text{ or } \cdots \text{ ad inf.}\} \to 0 \text{ as } n \to \infty.$$

If we denote by A_n the event $|X_n - X| > \epsilon$, we see that we are concerned here with

$P = P$ (realization of infinitely many events A_n) $= \lim\limits_{n\to\infty} \lim\limits_{\nu\to\infty} P(A_{n+1} \cup \cdots \cup A_{n+\nu})$[8] From Boole's inequality

$$P(A_{n+1} \cup \cdots \cup A_{n+\nu}) \leqq \sum_{k=n+1}^{n+\nu} P(A_k)$$

follows, at once, the fundamental BOREL-CANTELLI LEMMA. *If* $\sum_n P(A_n) < \infty$ *then* $P = 0$. This lemma can be extended, using sharper inequalities (Loève [32]).

[8] Already Poincaré considered such probabilities in his investigation of "recurrence" and this, before the notion of completely additive measures was born.

Now apply the Tchebicheff-Markov inequality

$$P\{|X_n - X| > \epsilon\} \leqq \frac{E|X_n - X|^r}{\epsilon^r}, \qquad r > 0,$$

and the Cantelli criterion follows: *if for some* $r > 0$, $\sum E|X_n - X|^r < \infty$ *then* $X_n \xrightarrow{\text{a.s.}} X$.

Applying it, with $r = 4$, to the Bernoulli case, Cantelli [6] obtained an almost immediate proof of Borel's result. An even simpler proof is as follows: $\sum_n E|f_{n^2} - p|^2 < \infty$ since $E(f_n - p)^2 = \dfrac{pq}{n}$, hence $f_{n^2} - p \xrightarrow{\text{a.s.}} 0$. Moreover, $|f_\nu - f_{n^2}| \leqq \dfrac{2}{n}$ for $0 \leqq \nu - n^2 \leqq 2n$, hence $f_\nu - f_{n^2} \to 0$ in the usual sense, uniformly in ν, and the theorem is proved This last method applies as well to sequences of dependent events $\{B_n\}$, which constitute a natural extension of the Bernoulli case. Let

$$p_1(n) = \frac{1}{n}\sum_{k=1}^{n} P(B_k), \qquad p_2(n) = \frac{1}{C_n^2}\sum_{1\leq k<l\leq n} P(B_k B_l),$$

$\delta_n = p_2(n) - p_1^2(n)$ (in the Bernoulli case $\delta_n = 0$!). It is very easy to show that $f_n - p_1(n) \to 0$ in probability if, and only if, $\delta_n \to 0$; this extends the Bernoulli theorem. Moreover, if $n|\delta_n| \leqq C < \infty$ then $f_n - p_1(n) \xrightarrow{\text{a.s.}} 0$ (Loève [31]), and Dvoretzky [10] proved that it is enough to have $\Sigma \dfrac{|\delta_n|}{n} < \infty$. Thus we have a simple extension of Borel's result.

The method used by Borel, while uselessly complicated in view of the result obtained, is very powerful and, by sharpening it, *the law of the iterated logarithm* (Khintchine [18]) follows.

Second Strong Limit Theorem. *In the Bernoulli case*

$$P\left\{\limsup_n \frac{S_n - ES_n}{\sigma_n(2\log\log\sigma_n)^{1/2}} = 1\right\} = 1.$$

where $\sigma_n = \sigma(S_n)$.

Let us use the following terminology (P. Lévy [26]). A non-decreasing sequence $\{\phi_n\}$ of positive numbers belongs to the *lower class* L, if the probability that $S_n \leqq \phi_n$, from some n onwards, is 1, and it belongs to the upper class U if this probability is 0. The following criterion (Kolmogorov) applies: *In the Bernoulli case* $\{\phi_n\}$ *belongs to L or U, respectively, according as* $\sum_n \dfrac{1}{\sigma_n^2}\phi_n e^{-1/2\phi_n^2} = \infty$ *or* $< \infty$. Clearly this result contains the Khintchine's LIT.

7. The general case. The question of domains of validity of the obtained results arises immediately and thus the SCLP appears in its present form. Let $S_n = \sum_{k=1}^{n} X_k$ be sums of r.v.'s X_k, independent or not. Find conditions for 1° a.s. convergence of $\dfrac{S_n}{n}$ or, more generally [31] of $\dfrac{S_n}{a_n}$, $a_n \uparrow \infty$ (SLLN). 2° the law

of the iterated logarithm (LIT) and, more generally, criteria for classifying sequences $\{\phi_n\}$.

The second problem, in the case of independent summands possesses almost complete solutions due, respectively, to Kolmogorov [17] and to Feller [13].

a. *If* $\sup |X_k| = o(\sigma_n/(\log\log\sigma_n)^{-1/2})$ *for* $k \leqq n$, *then LIT holds.*

b *If* $\sup |X_k| = O(\sigma_n/(\log\log\sigma_n)^{-3/2})$ *for* $k \leqq n$, *then the criterion for the Bernoulli case continues to hold.* (Feller also gave sharper criteria).

In the case of dependent summands general results were obtained by P. Lévy [26] and for Markov chains by Doeblin [7] The problem belongs (at present) to the domain of NC; it is complicated and pries deeply into the behavior of probabilities as $n \to \infty$. Yet, in the case of independence, the dichotomy into classes L and U is more general as shown by the following property (P. Levy [26]). *If* $\{S_n\}$ *is a sequence of consecutive sums of independent r.v's, and cannot be reduced by adding constants to an a s convergent sequence, then, for any given sequence* $\{c_n\}$ *of sure numbers,* $P(S_n > c_n$ *for an infinity of values of* $n) = 0$ *or* 1.

The SLLN problem *seems* easier. Nevertheless it is far from being solved; we don't even know necessary *and* sufficient conditions for the SLLN in the case of independent summands in terms of individual d f's.[9] The essential tools are, besides the fundamental Borel-Cantelli lemma, 1° the truncation method together with the convergence in r-mean: $X_n \xrightarrow{r} X$ if $E|X_n - X|^r \to 0 (r > 0)$, 2° the Kronecker lemma: *If* $\sum_n x_k/a_k$ *is convergent, then* $\frac{1}{a_n}\sum_{k=1}^n x_k \to 0$ $(a_n \uparrow \infty)$. It provides a possibility of transforming problems about the SLLN into those of a s. convergence of series of r.v 's, at least when sufficient conditions are sought for

In the case of independent summands one can start with the following property of series (Lévy [23]): *a.s. convergence of* $\sum_1^n X_k$ *is equivalent to convergence in probability.* (It can be shown that this property holds also for certain classes of dependent summands.) On the other hand, convergence in q m. $(r = 2)$ entails convergence in probability. Hence, when $EX_k^2 < \infty$, taking EX_k as the origin of values of X_k, it follows that *If* $\sum_n \sigma^2(X_n) < \infty$, *then* S_n *a.s converges.* Kolmogorov proved this result using his celebrated inequality which considerably strengthens that of Tchebicheff:

$$P\{\max_{k\leqq n}|S_n| > \epsilon\} \leqq \frac{\sigma^2(S_n)}{\epsilon^2}.$$

This inequality has been extended by P. Lévy [26], and by Loève [32] to dependent summands and conditions for a.s. convergence were deduced from it. If the EX_k^2 are not finite, the truncation method is applied Put $X_k' = X_k$, if $|X_k| \leqq 1$ and $= 0$ if $|X_k| > 1$. Then (Khintchine-Kolmogorov) $\sum_n X_n$,

[9] A first step in this direction is due to U. V. Prokhorov, "On the strong law of large numbers" (in Russian), *Dokl Ak. Nauk* Vol 69 (1949), pp 607–610 See also a paper by K. L. Chung to appear in the *Proceedings of the Second Berkeley Symposium.*

where X_n are independent r.v.'s, is a.s. convergent if, and only if, $\sum_n P(X_n \neq X'_n)$, $\sum_n \sigma^2(X'_n)$, $\sum_n (X'_n)$ *converge*

It is not difficult to obtain conditions for series of dependent summands Let $q_n(t) = P\{ | X_n | > t\}$, $\xi_n = \int_{-\epsilon}^{+\epsilon} x\, dF'_n(x)$, where $F'_n(n)$ is the conditional d.f. of X_n, given $X_1, \cdots X_{n-1}$. *If* $\sum_n \int_0^\epsilon t q_n(t)\, dt < \infty$ *for an* $\epsilon > 0$, *then* $\sum_n (X_n - \xi_n)$ *a.s. converges.*

By using Kronecker's lemma the results above yield immediately sufficient conditions for the SLLN. Those which come from the last one would in turn yield without difficulty the following: *Let* $a_n \uparrow \infty$ *and* $\eta_n = \int_{-\epsilon a_n}^{+\epsilon a_n} x\, dF'_n(x)$. *If* $\sum_n q_n(a_n t) \leqq q(t)$ *and* $\int_0^\epsilon t q(t)\, dt < \infty$, *then* $\frac{1}{a_n} \sum_{k=1}^{n} (X_k - \eta_k) \to 0$.

Take now the particular case: $a_n = n^r$, and X_k's *independent and identically distributed.* From the stated result follows:

1. *If* $EX_k = m$ *exist, then* $\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{\text{a.s}} m$ *and conversely* (Kolmogorov).

2. *If* $0 < r < 2$, $r \neq 1$, $E | X_k |^r < \infty$ *and* $\lim_{a \to \infty} \int_{-a}^{+a} x\, dF_n(x) = 0$, *then*

$$\frac{1}{n^{1/r}} \sum_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} 0 \text{ (Marcinkiewicz).}$$

Other conditions for SLLN, in the case of dependence, are known (Lévy [27], Loève [32]).

The above result of Kolmogorov is a particular case of the celebrated ergodic theorem (Birkhoff [3]) which can be considered as a SLLN for a special case of dependence Let A_n be an event defined on the set $\{X_{k_1}, \cdots X_{k_n}\}$ and let $A_n^{(m)}$ be an event defined in the same manner on the translated set $\{X_{k_1+m}, \cdot \cdot, X_{k_n+m}\}$. The sequence $\{X_k\}$ is called *stationary* if $P(A_n^{(m)}) = P(A_n)$ for every finite set $\{k_1, \cdots, k_n\}$ and every finite m. The ergodic theorem states that *If the sequence* $\{X_k\}$ *is stationary and* $E\, |X_k| < \infty$, *then* $\frac{1}{n} \sum_{k=1}^{n} X_k$ *converges a.s.*[10]

However an unsatisfactory feature of Birkhoff's theorem (and of its extensions) is that the conditions are not asymptotic—they have to be satisfied for every n and not for $n \to \infty$—while the conclusion is an asymptotic one. Let us only mention that more satisfactory ones, at least from this point of view, which contain the previous ones, can be found.

[10] For about fifteen years Khintchine, Kolmogorov, Wiener, Yosida and Kakutani, F. Riesz, worked to simplify the proof of this theorem. It is only lately that its domain of validity has been extended by Hurewicz, by Halmos, and by Dunford and Miller See also a forthcoming paper by the author in the *Proceedings of the Second Berkeley Symposium.*

The bird's-eye view above of the SCLP shows that this problem is only in a tentative stage, perhaps because no adequately powerful methods or no adequately general approach to the problem had been found until now.

REFERENCES

[1] G. Bawly, Über einige Verallgemeinerungen der Grenzwertsätze der Wahrscheinlichkeitsrechnung," *Rec. Math. Moscow*, Vol 43 (1936), pp 917–929

[2] J. Bernoulli, *Ars conjectandi*, 1713.

[3] S. Bernstein, "Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes," *Math Ann.*, Vol. 97 (1927), pp. 1–59

[4] G D Birkhoff, "Proof of the ergodic theorem," *Proc Nat. Acad Sc.* Vol 17 (1931), pp. 650–665.

[5] E. Borel, "Sur les probabilités dénombrables et leurs applications arithmétiques," *Circ. Mat. d Palermo*, Vol. 26 (1909), pp 247–271

[6] F P. Cantelli, "Sulla probabilita come limite della frequenza," *Rend Accad. d. Lincei*, Vol 26 (1917), pp 39–45.

[7] A. De Moivre, *Miscellanea analytica*, 1730.

[8] W. Doeblin, "Sur les propriétés asymptotiques . . . de chaînes simples," *Bull Math. Soc Roum Sc* (1937), pp. 1–120

[9] W. Doeblin, "Sur les sommes d'un grand nombre de variables aléatoires indépendantes," *Bull. Sc. Math.*, Vol 63 (1939), pp. 23–64.

[10] A. Dvoretzky, "On the strong stability of a sequence of events, *Annals of Math. Stat* , Vol. 20 (1949), pp 296–299.

[11] W Feller, "Ueber den Zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung," *Math. Zeit.*, Vol 40 (1935), pp. 521–539; Vol. 42 (1937), pp 301–312.

[12] W Feller, "Ueber das Gesetz der grossen Zahlen," *Acta. Litt Ac Scient.* Vol. 8 (1937), pp. 191–201.

[13] W. Feller, "The general form of the so-called law of the iterated logarithm," *Trans Am Math. Soc.*, Vol 54 (1943), pp. 373–402.

[14] B. Gnedenko, "On convergence of distribution laws of independent random variables" (in Russian), *Dokl Ak. Nauk*, Vol. 18 (1938), pp. 231–234

[15] B. Gnedenko, "On the theory of limit theorems for sums of independent random variables" (in Russian), *Bull Ac Sci. URSS* (1939), pp. 181–232, pp 643–647

[16] A Kolmogoroff, "Ueber die Summen durch den Zufall bestimmter unabhangiger Grössen," *Math Ann* , Vol. 99 (1928), pp 309–319, 102 (1929), pp. 484–489

[17] A. Kolmogoroff, "Ueber das Gesetz des iterierten Logarithms," *Math. Ann* , Vol. 101 (1929), pp 126–135

[18] A Kolmogoroff, "Sulla forma generale di un processo stocastico omogeneo," *Rend Accad d Lincei*, Vol. 15 (1932), pp. 805–808, pp 866–869

[19] A Khintchine, "Ueber einen Satz der Wahrscheinlichkeitsrechnung," *Fund Math.*, Vol 6 (1924), pp. 9–20

[20] A Khintchine, "Zur Theorie der unbeschrankt teilbaren Verteilungsgesetze," *Rec Math. Moscow*, Vol 44 (1937), pp 79–119.

[21] P S Laplace, *Traité des Probabilities*, 1812.

[22] P Lévy, *Calcul des Probabilités*, Gauthier-Villars, 1925

[23] P Lévy, "Sur les séries dont les termes sont des variables éventuelles indépendantes," *Studia Math* Vol 3 (1931), pp 117–155.

[24] P Lévy, "Sur les integrales dont les éléments sont des variables aléatoires indépendantes," *Ann d Scuola Nor di Pisa*, B (1934), pp 337–366

[25] P. Lévy, "Propriétés asymptotiques des sommes de variables aléatoires indépendantes on enchaînées," *Jour. Math. Pures Appl* , Vol 14 (1935), pp. 347–402

[26] P. Lévy, "La loi forte des grands nombres pour les variables aléatoires enchaînées," *Jour. Math Pures Appl*, Vol. 15 (1936), pp 11–24
[27] P Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villers, 1937
[28] A. Liapounoff, "Sur une proposition de la théorie des probabilités," *Bull. Acad Sci. St. Petersbourg*, (5); Vol. 13 (1900), pp 359–386
[29] A Liapounoff, "Nouvelle forme du théorème sur la limite de probabilités," *Mem Ac. St. Petersbourg*, (8) Vol 12 (1901).
[30] J. Lindeberg, "Eine nene Heileitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung," Vol 15 (1922), pp 211–225
[31] M Loève, "La tendance centrale pour des variables aléatoires liées," *C. R. Ac. Sc. Paris*, Vol 212 (1940).
[32] M. Loève, "Etude asymptotique des sommes de variables aléatoires liées," *Jour. Math Pures Appl.*, Vol. 24 (1945), pp 249–318
[33] M Loève, "Sur l'équivalence asymptotique des lois," *C. R Ac. Sc. Paris*, Vol. 227 (1948), pp. 1335–1337.
[34] M. Loève, "On the Central Probability problem," *Proc Nat. Acad Sci.*, Vol 35 (1949), pp. 328–332.
[35] M Loève, "On sets of Probability laws and their limit elements," *Statistical Series*, Vol. 1, pp. 53–88, University of California Press, Berkeley, California
[36] J Marcinkiewicz, "Sur les fonctions indépendantes," *Fund Math.*, Vol 30 (1938), pp 349–364.
[37] A Markoff, *Calculus of probabilities* (Russian), 1913.
[38] Poisson, *Recherches sur la Probabilité des Jugements* 1837, pp. 205–207.
[39] D Raïkoff, "On connection between the central limit theorem of probability theory and the law of large numbers," *Bull. Ac Sci URSS*, Ser. Math , (1938), pp. 323–338
[40] P. L. Tchebicheff, *Oeuvres*, Vol I, pp. 687–694; Vol II, pp. 481–492.

A RANDOM VARIABLE RELATED TO THE SPACING OF SAMPLE VALUES

By B. Sherman[1]

University of Southern California

1. Introduction and summary. Let x be a random variable with continuous distribution function $F(x)$. Then $y = F(x)$ is a random variable uniformly distributed over [0, 1]. If $x_1, x_2, \cdots, x_n$ is an ordered sample of n values from the population $F(x)$ then $y_1, y_2, \cdots, y_n$ $(y_i = F(x_i))$ is an ordered sample of n values from a uniform distribution over [0, 1]. For n large it is reasonable to expect that the y_i should be fairly uniformly spaced. Measures of the deviation from uniform spacing can be devised in various ways. Thus Kimball [2] has studied the random variable

$$\alpha = \sum_{i=1}^{n+1} \left(F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right)^2,$$

where $x_0 = -\infty$ and $x_{n+1} = +\infty$, conjecturing that $\alpha^{\frac{1}{2}}$ is asymptotically normally distributed. Moran [3] has studied the random variable

$$\beta = \sum_{i=1}^{n+1} (F(x_i) - F(x_{i-1}))^2,$$

which differs from α only by the quantity $-2/(n+1) + (n+1)^{-2}$, and has proved that β is asymptotically normally distributed. Somewhat related to these two random variables is the quantity ω^2 introduced by Smirnoff [4]. This is

$$\omega^2 = n \int_{-\infty}^{\infty} (F(x) - F^*(x))^2 \, dF(x),$$

although it is slightly more generally defined in Smirnoff's paper. Here $F^*(x)$ is the sample distribution function ([1], page 325) of a sample of n values from the population with continuous distribution function $F(x)$. The variable ω^2 may be written ([1], page 451)

$$\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(F(x_i) - \frac{2i-1}{2n}\right)^2.$$

$(2i - 1)/2n$ is the midpoint of the interval $((i-1)/n, i/n)$. Thus, if [0, 1] is partitioned into n equal subintervals then ω^2 measures the deviation of the sample values $y_i = F(x_i)$, $i = 1, 2, \cdots, n$, from the midpoints of these intervals. Smirnoff has investigated the asymptotic behavior of ω^2 obtaining a rather complicated non-normal asymptotic distribution.

[1] I wish to thank Professors J. W. Tukey and S. S. Wilks for their helpful suggestion and criticism.

It is possible to construct a definition of deviation from uniform spacing which permits a broader investigation than these random variables. This is

$$\omega_n = \frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right|,$$

where again $x_0 = -\infty$ and $x_{n+1} = +\infty$ and $F(x)$ is a continuous distribution function. (In Theorems 3 and 4 it is assumed additionally that $F'(x)$ exists and is continuous except for a finite number of points). It is to be noted that

$$0 \leqq \omega_n \leqq 1.$$

Generally speaking use of the absolute value in circumstances like this is an undesirable procedure, but it turns out that ω_n is relatively easy to handle, allowing a fairly simple calculation of its moments (which are independent of $F(x)$). These are ($\mu = \min(k, n)$)

$$\alpha_{nk} = E(\omega_n^k) = \binom{n+k}{k}^{-1} \sum_{s=0}^{\mu-1} \binom{n+1}{s+1} \binom{k-1}{s} \left(\frac{n-s}{n+1}\right)^{n+k}.$$

Thus in particular the mean of ω_n is

$$E(\omega_n) = \left(\frac{n}{n+1}\right)^{n+1} \to \frac{1}{e},$$

and the variance is

$$D^2(\omega_n) = E(\omega_n^2) - E^2(\omega_n) = \frac{2n^{n+2} + n(n-1)^{n+2}}{(n+2)(n+1)^{n+2}} - \left(\frac{n}{n+1}\right)^{2n+2}$$

$$\sim \frac{2e-5}{e^2} \frac{1}{n}.$$

These results will be established in Theorem 1. From the moments the characteristic function of ω_n may be obtained, and indeed in finite terms From the characteristic function the distribution function of ω_n may be readily calculated. The distribution function is written out explicitly at the end of Theorem 1.

To determine the asymptotic distribution of the standardized variable

$$\frac{\omega_n - E(\omega_n)}{D(\omega_n)},$$

it is sufficient to examine the behaviour as $n \to \infty$ of the moments of this variable or equivalently the moments of the variable

$$\left(\frac{ne^2}{2e-5}\right)^{1/2} \left(\omega_n - \frac{1}{e}\right).$$

For it is easy to show that if the moments of the standardized variable approach the moments of a unique distribution function $F(x)$ then the distribution function of the standardized variable approaches $F(x)$. In this manner it is proved

in Theorem 2 that the distribution function of the standardized variable approaches normality.

Since the asymptotic distribution of the standardized variable

$$\frac{\omega_n - E(\omega_n)}{D(\omega_n)}$$

is known it may be used as a test for goodness of fit if the number of sample values is large. Thus suppose $x_1, x_2, \cdots, x_n$ is an ordered sample of n values from some population and we wish to test the hypothesis that the population has the distribution function $F(x)$. Then we calculate the quantity

$$\left| \frac{1}{D(\omega_n)} \left[\frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| - E(\omega_n) \right] \right| = X_n,$$

and if this quantity exceeds a certain value which depends on the level of significance at which we are working we reject the hypothesis. Let us say that $P(X_n > A) = B$. The probability of rejecting the hypothesis when it is indeed true is then precisely B and this is small if A is sufficiently large. But suppose that the hypothesis is false and the sample values come from a population whose distribution function $G(x) \not\equiv F(x)$ Then we would desire the following property to hold for the random variable X_n, namely, for any fixed positive A the probability that X_n exceeds A approaches 1 as $n \to \infty$. For in this case (and when n is large) we are almost certain to reject the null hypothesis when it is false. A test for goodness of fit which satisfies this criterion, i.e where the probability of rejection approaches 1 as $n \to \infty$ when the null hypothesis is false, is called consistent by Wald and Wolfowitz [5]. We wish to prove then that the test for goodness of fit which uses the random variable X_n is consistent. To express the matter formally we wish to prove that (the probability density element of $x_1, x_2, \cdots, x_n$ is $n!\, dG(x_1)\, dG(x_2) \cdots dG(x_n)$ in the region

$$-\infty < x_1 < x_2 < \cdots < x_n < +\infty$$

and zero outside that region).

$$\lim_{n\to\infty} \int_{D_1} \cdots \int dG(x_1) \cdots dG(x_n) = \begin{cases} \dfrac{2}{\sqrt{2\pi}} \displaystyle\int_A^\infty e^{-(x^2/2)}\, dx & \text{if } F(x) \equiv G(x), \\ 1 & \text{if } F(x) \not\equiv G(x), \end{cases}$$

where D_1 is the domain

$$-\infty < x_1 < x_2 < \cdots < x_n < +\infty,$$

$$\left| \frac{1}{D(\omega_n)} \left[\frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| - E(\omega_n) \right] \right| > A$$

The first assertion here is proved in Theorem 2. The second assertion is equivalent to proving that for any fixed positive A

$$\lim_{n\to\infty} \int_{D_2} \cdots \int dG(x_1)\, dG(x_2) \cdots dG(x_n) = 0, \tag{0.1}$$

where D_2 is the domain

$$-\infty < x_1 < x_2 < \cdots < x_n < +\infty,$$

$$E(\omega_n) - AD(\omega_n) < \frac{1}{2}\sum_{i=1}^{n+1}\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right| < E(\omega_n) + AD(\omega_n),$$

when $F(x) \not\equiv G(x)$. Now $D(\omega_n)$ is of order $n^{-1/2}$, $E(\omega_n) = e^{-1}+$ terms of order n^{-1} and A is fixed. Hence it is sufficient to show that, if $x_1, x_2, \cdots, x_n$ is an ordered sample of n values from a population with distribution function $G(x)$, then the random variable

$$\Omega_n = \frac{1}{2}\sum_{i=1}^{n+1}\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|$$

(it is necessary to draw a distinction between ω_n and Ω_n since $F(x) \not\equiv G(x)$) has a mean $L_n \to L \neq e^{-1}$ and a variance $D^2(\Omega_n) \to 0$. For then we have, when n is large enough so that the interval

$$[E(\omega_n) - AD(\omega_n), \qquad E(\omega_n) + AD(\omega_n)]$$

falls outside $[L - \frac{1}{2}\,|\,L - e^{-1}\,|, L + \frac{1}{2}\,|\,L - e^{-1}\,|]$ and $|\,L_n - L\,| < \frac{1}{4}\,|\,L - e^{-1}\,|$,

$$\begin{aligned} P(E(\omega_n) - AD(\omega_n) < \Omega_n < E(\omega_n) + AD(\omega_n)) \\ &\leqq P(|\,\Omega_n - L\,| \geqq \tfrac{1}{2}\,|\,L - e^{-1}\,|) \\ &\leqq P(|\,\Omega_n - L_n\,| \geqq \tfrac{1}{4}\,|\,L - e^{-1}\,|) \\ &\leqq \frac{E(|\,\Omega_n - L_n\,|)}{\frac{1}{4}\,|\,L - e^{-1}\,|} \leqq \frac{D(\Omega_n)}{\frac{1}{4}\,|\,L - e^{-1}\,|}, \end{aligned}$$

and this implies (0.1).

But now in Theorem 3 it is shown that the mean of the random variable Ω_n is (writing $k(x) = GF^{-1}(x)$, $k(x)$ a monotonic function such that $k(0) = 0$ and $k(1) = 1$)

$$\int_0^{n/n+1}\left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n dx.$$

This expression approaches

$$\int_0^1 e^{-k'(x)}\, dx$$

and this integral can assume the value e^{-1}, which is its minimum relative to the class of monotonic functions such that $k(0) = 0$ and $k(1) = 1$, only when $k(x) \equiv x$ i.e. $F(x) \equiv G(x)$ Finally in Theorem 4 we prove that $D^2(\Omega_n) \to 0$ and thus it is established that the test for goodness of fit based on X_n is consistent.

2. Moments and asymptotic distribution of ω_n.

THEOREM 1. *Let $F(x)$ be a continuous distribution function If $x_1, x_2, \cdots, x_n$ is an ordered sample of n values from the population whose distribution function is $F(x)$ then the random variable*

$$\omega_n = \frac{1}{2}\sum_{i=1}^{n+1}\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|,$$

where $x_0 = -\infty$ *and* $x_{n+1} = +\infty$, *has the moments*

$$\alpha_{nk} = E(\omega_n^k) = \binom{n+k}{k}^{-1} \sum_{s=0}^{\mu-1} \binom{n+1}{s+1} \binom{k-1}{s} \left(\frac{n-s}{n+1}\right)^{n+k},$$

where $\mu = \min(k, n)$.

The probability density element of the x_i is ([6], page 90)

$$n!\, dF(x_1)\, dF(x_2) \cdots dF(x_n)$$

in the domain D_x: $-\infty < x_1 < x_2 < \cdots < x_n < +\infty$ and zero outside of this domain. Then

$$\alpha_{nk} = n! \int \cdots \int_{D_x} \omega_n^k \, dF(x_1)\, dF(x_2) \cdots dF(x_n).$$

If we make the transformation $y_i = F(x_i)$, $\quad i = 1, 2, \cdots, n$, then

$$\alpha_{nk} = n! \int \cdots \int_{D_y} \left[\frac{1}{2} \sum_{i=1}^{n+1} \left| y_i - y_{i-1} - \frac{1}{n+1} \right| \right]^k dy_1\, dy_2 \cdots dy_n,$$

where D_y is the domain $0 < y_1 < y_2 < \cdots < y_n < 1$, thus indicating that the moments of ω_n (and therefore also the distribution function of ω_n) are independent of $F(x)$. Here $y_0 = 0$ and $y_{n+1} = 1$. The transformation

$$\begin{array}{ll} u_1 = y_1, & y_1 = u_1, \\ u_2 = y_2 - y_1, & y_2 = u_1 + u_2, \\ \cdots & \cdots \\ u_n = y_n - y_{n-1}, & y_n = u_1 + u_2 + \cdots + u_n, \\ u_{n+1} = y_{n+1} - y_n, & y_{n+1} = u_1 + u_2 + \cdots + u_n + u_{n+1} = 1, \end{array}$$

whose Jacobian is 1, then yields

$$\begin{aligned} \alpha_{nk} &= n! \int \cdots \int_{D_u} \left[\frac{1}{2} \sum_{i=1}^{n+1} \left| u_i - \frac{1}{n+1} \right| \right]^k du_1\, du_2 \cdots du_n \\ &= n! \int \cdots \int_{D_u} \left[\frac{1}{2} \sum_{i=1}^{n} \left| u_i - \frac{1}{n+1} \right| \right. \\ &\qquad\qquad \left. + \frac{1}{2} \left| \frac{n}{n+1} - (u_1 + u_2 + \cdots + u_n) \right| \right]^k du_1 \cdots du_n, \end{aligned}$$

where D_u is the domain $\sum_{i=1}^{n} u_i < 1$, $u_i > 0$, $i = 1, 2, \cdots, n$.

The domain D_u can be regarded as the union of $2^{n+1}-2$ subdomains in the following way First the hyperplane $u_1 + u_2 + \cdots + u_n = n/(n+1)$ divides the domain into two parts. In the part of the domain below the hyperplane, i e where $u_1 + u_2 + \cdots + u_n < n/(n+1)$, we have a subdomain defined by the statement: k of the variables u_i are greater than $(n+1)^{-1}$ and the

residual group of $n - k$ u_i are less than $(n + 1)^{-1}$. There are $\binom{n}{k}$ such subdomains and it is clear that, because of the symmetry in the u_i, the intregal of $\left[\frac{1}{2}\sum_{i=1}^{n+1}\left|u_i - \frac{1}{n+1}\right|\right]^k$ over each such subdomain is the same There are altogether $\sum_{k=0}^{n-1}\binom{n}{k} = 2^n - 1$ such subdomains. $k \neq n$ because of the inequality $u_1 + u_2 + \cdots + u_n < n/(n+1)$. In the part of the domain above the hyperplane

$$u_1 + u_2 + \cdots + u_n = n/(n + 1),$$

i.e. where $u_1 + u_2 + \cdots + u_n > n/(n + 1)$, the reasoning is exactly the same except that here $k \neq 0$. Thus we may write

$$\alpha_{nk} = n! \sum_{r=0}^{n-1}\binom{n}{r}\int\cdots\int_{D_{r1}}\left[\sum_{i=r+1}^{n}\left(\frac{1}{n+1} - u_i\right)\right]^k du_1\, du_2 \cdots du_n$$
$$+ n! \sum_{r=1}^{n}\binom{n}{r}\int\cdots\int_{D_{r2}}\left[\sum_{i=1}^{r}\left(u_i - \frac{1}{n+1}\right)\right]^k du_1\, du_2 \cdots du_n,$$

where D_{r1} is the domain

$$\sum_{i=1}^{n} u_i < \frac{n}{n+1}, \qquad u_i > \frac{1}{n+1} \qquad (i = 1, 2, \cdots, r),$$
$$0 < u_i < \frac{1}{n+1} \qquad (i = r + 1, \cdots, n),$$

and D_{r2} is the domain

$$\frac{n}{n+1} < \sum_{i=1}^{n} u_i < 1, \qquad u_i > \frac{1}{n+1} \qquad (i = 1, 2, \cdots, r),$$
$$0 < u_i < \frac{1}{n+1} \qquad (i = r + 1, \cdots, n).$$

If we introduce the variables

$$z_i = u_i - \frac{1}{n+1} \qquad (i = 1, 2, \cdots, r),$$
$$z_i = \frac{1}{n+1} - u_i \qquad (i = r + 1, \cdots, n),$$

we get

$$\alpha_{nk} = n! \sum_{r=0}^{n-1}\binom{n}{r}\int\cdots\int_{\Delta_{r1}}\left(\sum_{i=r+1}^{n} z_i\right)^k dz_1 \cdots dz_n$$
$$+ n! \sum_{r=1}^{n}\binom{n}{r}\int\cdots\int_{\Delta_{r2}}\left(\sum_{i=1}^{r} z_i\right)^k dz_1 \cdots dz_n,$$

where Δ_{r1} is the domain

$$\sum_{i=1}^{r} z_i < \sum_{i=r+1}^{n} z_i, \qquad z_i > 0 \qquad (i = 1, 2, \cdots, r),$$

$$\frac{1}{n+1} > z_i > 0 \qquad (i = r+1, \cdots, n),$$

and Δ_{r2} is the domain

$$\sum_{i=r+1}^{n} z_i < \sum_{i=1}^{r} z_i < \frac{1}{n+1} + \sum_{i=r+1}^{n} z_i, \qquad z_i > 0 \qquad (i = 1, 2, \cdots, r),$$

$$\frac{1}{n+1} > z_i > 0 \qquad (i = r+1, \cdots, n).$$

To effect the integrations with respect to the variables $z_1, z_2, \cdots z_r$ we take as volume element in the r-space of $z_1, z_2, \cdots z_r$ the volume between the hyperplanes $z_1 + z_2 + \cdots + z_r = C$, $z_i > 0$ and $z_1 + z_2 + \cdots + z_r = C + dC$, $z_i > 0$. This volume element is $d\,\frac{C^r}{r!} = \frac{C^{r-1}}{(r-1)!}\,dC$. Thus

$$\alpha_{nk} = n! \sum_{r=0}^{n-1} \binom{n}{r} \int_0^{1/n+1} \cdots \int_0^{1/n+1} \left[\int_0^{\sum_{i=r+1}^{n} z_i} \frac{C^{r-1}}{(r-1)!}\,dC\right] \left(\sum_{i=r+1}^{n} z_i\right)^k dz_{r+1} \cdots dz_n$$

$$+ n! \sum_{r=1}^{n} \binom{n}{r} \int_0^{1/n+1} \cdots \int_0^{1/n+1} \left[\int_{\sum_{i=r+1}^{n} z_i}^{(1/n+1)+\sum_{i=r+1}^{n} z_i} \frac{C^{k+r-1}}{(r-1)!}\,dC\right] dz_{r+1} \cdots dz_n$$

$$= n! \sum_{r=0}^{n-1} \binom{n}{r} \int_0^{1/n+1} \cdots \int_0^{1/n+1} \frac{1}{r!} (z_{r+1} + \cdots + z_n)^{k+r}\, dz_{r+1} \cdots dz_n$$

$$+ n! \sum_{r=1}^{n} \binom{n}{r} \int_0^{1/n+1} \cdots \int_0^{1/n+1} \frac{1}{(k+r)(r-1)!} \cdot \left(\frac{1}{n+1} + z_{r+1} + \cdots + z_n\right)^{k+r} dz_{r+1} \cdots dz_n$$

$$- n! \sum_{r=1}^{n} \binom{n}{r} \int_0^{1/n+1} \cdots \int_0^{1/n+1} \frac{1}{(k+r)(r-1)!} \cdot (z_{r+1} + \cdots + z_n)^{k+r}\, dz_{r+1} \cdots dz_n.$$

In order to perform these integrations we use the formula

$$\int_0^A \cdots \int_0^A (B + x_1 + x_2 + \cdots + x_n)^m\, dx_1 \cdots dx_n = \frac{m!}{(m+n)!} \sum_{q=0}^{n} (-1)^{n-q} \binom{n}{q} (B + qA)^{m+n},$$

which is established immediately by induction on n. Then

$$\alpha_{nk} = n! \sum_{r=0}^{n-1} \sum_{q=0}^{n-r} \frac{(-1)^{n-r-q}}{r!} \frac{(k+r)!}{(n+k)!} \binom{n}{r} \binom{n-r}{q} \left(\frac{q}{n+1}\right)^{n+k}$$

$$+ n! \sum_{r=1}^{n} \sum_{q=0}^{n-r} \frac{(-1)^{n-r-q}}{(r-1)!} \frac{(k+r-1)!}{(n+k)!} \binom{n}{r} \binom{n-r}{q} \left(\frac{1+q}{n+1}\right)^{n+k}$$

$$- n! \sum_{r=1}^{n} \sum_{q=0}^{n-r} \frac{(-1)^{n-r-q}}{(r-1)!} \frac{(k+r-1)!}{(n+k)!} \binom{n}{r} \binom{n-r}{q} \left(\frac{q}{n+1}\right)^{n+k}.$$

The first of these double sums is equal to

$$\frac{n!\,k!}{(n+k)!} \sum_{q=1}^{n} \sum_{r=0}^{n-q} (-1)^{n-r-q} \binom{n}{q} \binom{n-q}{r} \binom{k+r}{k} \left(\frac{q}{n+1}\right)^{n+k}$$

$$= \binom{n+k}{k}^{-1} \sum_{q=1}^{n} \binom{n}{q} \left(\frac{q}{n+1}\right)^{n+k} \left[\sum_{r=0}^{n-q} (-1)^{n-r-q} \binom{n-q}{r} \binom{k+r}{k}\right].$$

Let us assume first that $n \geqq k$. The expression within the brackets is the coefficient of x^{n-q} in $(1-x)^{n-q}(1/(1-x)^{k+1}) = (1-x)^{n-q-k-1}$ and this is $\neq 0$ only when $q \geqq n-k$ and then it has the value $\binom{k}{n-q}$. Thus the first double sum is equal to

$$\binom{n+k}{k}^{-1} \sum_{q=n-k}^{k} \binom{k}{n-q} \binom{n}{q} \left(\frac{q}{n+1}\right)^{n+k}$$

$$= \binom{n+k}{k}^{-1} \sum_{s=0}^{k} \binom{k}{s} \binom{n}{s} \left(\frac{n-s}{n+1}\right)^{n+k}.$$

Similarly the second double sum is equal to

$$\binom{n+k}{k}^{-1} \sum_{s=0}^{k-1} \binom{k-1}{s} \binom{n}{s+1} \left(\frac{n-s}{n+1}\right)^{n+k},$$

and the third is equal to

$$\binom{n+k}{k}^{-1} \sum_{s=1}^{k} \binom{k-1}{s-1} \binom{n}{s} \left(\frac{n-s}{n+1}\right)^{n+k}.$$

Thus, using the identity

$$\binom{k}{s}\binom{n}{s} + \binom{k-1}{s}\binom{n}{s+1} - \binom{k-1}{s-1}\binom{n}{s} = \binom{n+1}{s+1}\binom{k-1}{s},$$

we get

$$\alpha_{nk} = \binom{n+k}{k}^{-1} \sum_{s=0}^{k-1} \binom{n+1}{s+1} \binom{k-1}{s} \left(\frac{n-s}{n+1}\right)^{n+k}.$$

If however $k > n$ then a similar argument shows that we get an expression for α_{nk} which differs from the above only in the upper limit of the summation, which is $n - 1$ in this case. Thus the theorem is proved.

The distribution function of ω_n is

$$F(x) = 1 + \sum_{q=0}^{n-r-1} \sum_{p=0}^{q} (-1)^{q-p+1} \binom{n}{p} \binom{n+1}{q+1} \cdot \binom{n+q-p}{n} \left(\frac{n-q}{n+1}\right)^p \left(\frac{n-q}{n+1} - x\right)^{n-p},$$

where r is the non-negative integer determined by the inequality

$$\frac{r}{n+1} \leqq x < \frac{r+1}{n+1}.$$

$F(x) = 0$ when $x \leqq 0$, $F(x) = 1$ when $x \geqq n/(n+1)$ and $F(x)$ is a polynomial of degree n in each of the intervals

$$\left(\frac{i-1}{n+1}, \frac{i}{n+1}\right), \qquad i = 1, 2, \cdots n.$$

THEOREM 2. *The random variable ω_n is asymptotically normally distributed $(E(\omega_n), D(\omega_n))$; i.e., the distribution function of the standardized variable*

$$\frac{\omega_n - E(\omega_n)}{D(\omega_n)}$$

approaches

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t^2/2)}\, dt.$$

It is sufficient to prove that the moments of the standardized variable approach the moments of the normal distribution. For in general it is known that if the moments α_{nk} of $F_n(x)$ approach the moments α_k of a uniquely determined distribution function $F(x)$, then $F_n(x)$ converges to $F(x)$ in every continuity point of the latter (M. G Kendall, *Advanced Theory of Statistics*, Vol. 1, Third edition, Charles Griffin and Co., 1943, pp. 110–112).

Now $E(\omega_n) \to \frac{1}{e}$ and $D^2(\omega_n) \sim \frac{2e-5}{e^2} \frac{1}{n} = \frac{c}{n}$, so that the two variables $\frac{\omega_n - E(\omega_n)}{D(\omega_n)}$ and $\left(\frac{n}{c}\right)^{\frac{1}{2}}\left(\omega_n - \frac{1}{e}\right)$ have the same limiting distribution. Thus it is sufficient to prove that the moments of $\left(\frac{n}{c}\right)^{\frac{1}{2}}\left(\omega_n - \frac{1}{e}\right)$ tend to the moments of the normal distribution. In the following argument we take $\mu = k$ since $n \to \infty$.

$$E\left[\left(\frac{n}{c}\right)^{m/2}\left(\omega_n - \frac{1}{e}\right)^m\right] = \left(\frac{n}{c}\right)^{m/2} \sum_{k=0}^{m} \binom{m}{k} \alpha_k \left(-\frac{1}{e}\right)^{m-k}$$

$$(2.1) \qquad = \frac{n^{m/2}\, m!}{(2e-5)^{m/2}} \left[\frac{(-1)^m}{m!} + \sum_{k=1}^{m} \sum_{s=0}^{k-1} \frac{(-1)^{m-k}\, n!\, e^k}{(n+k)!\,(m-k)!} \cdot \binom{n+1}{s+1} \binom{k-1}{s} \left(\frac{n-s}{n+1}\right)^{n+k}\right].$$

Suppose now that it has been proved that $E\left[\left(\frac{n}{c}\right)^m\left(\omega_n - \frac{1}{e}\right)^{2m}\right]$ tends to a finite limit as $n \to \infty$, i.e., that the limiting moments of order $2m$ exist, $m = 1, 2, \cdots$. If m is odd

$$\left| E\left[\left(\frac{n}{c}\right)^{m/2}\left(\omega_n - \frac{1}{e}\right)^m\right]\right|$$

$$\leqq E\left[\left|\left(\frac{n}{c}\right)^{m/2}\left(\omega_n - \frac{1}{e}\right)^m\right|\right] \leqq \left\{E\left[\left(\frac{n}{c}\right)^m\left(\omega_n - \frac{1}{e}\right)^{2m}\right]\right\}^{1/2}.$$

Hence, if m is odd, $E\left[\left(\frac{n}{c}\right)^{m/2}\left(\omega_n - \frac{1}{e}\right)^m\right]$ is bounded as $n \to \infty$. Now the expression in the bracket on the right of (2.1) can be expanded in a convergent power series in n^{-1} provided that $n > m$. Because of the factor $n^{m/2}$ and because the left hand side of (2.1) is bounded as $n \to \infty$ this power series must have $\frac{a_p}{n^p}$, where $p \geqq \frac{m+1}{2}$ (since m is odd), as its initial non-vanishing term. But then the left hand side of (2.1) must approach 0 as $n \to \infty$. Thus if the limiting moments of even order exist the limiting moments of odd order are zero. We may now restrict the discussion to even order moments.

Replacing m by $2m$ in (2.1)

$$E\left[\left(\frac{n}{c}\right)^m\left(\omega_n - \frac{1}{e}\right)^{2m}\right] = \frac{n^m(2m)!}{(2e-5)^m}\left[\frac{1}{(2m)!}\right.$$

$$\left. + \sum_{k=1}^{2m}\sum_{s=0}^{k-1} \frac{(-1)^k n! e^k}{(n+k)!(2m-k)!}\binom{n+1}{s+1}\binom{k-1}{s}\left(\frac{n-s}{n+1}\right)^{n+k}\right].$$

Let us introduce the index $q = k - s - 1$ which runs from 0 to $2m - 1$. Then

$$E\left[\left(\frac{n}{c}\right)^m\left(\omega_n - \frac{1}{e}\right)^{2m}\right] = \frac{n^m(2m)!}{(2e-5)^m}\left[\frac{1}{(2m)!}\right.$$

$$\left. + \sum_{q=0}^{2m-1}\sum_{k=q+1}^{2m} \frac{(-1)^k n! e^k}{(n+k)!(2m-k)!}\binom{n+1}{k-q}\binom{k-1}{q}\left(\frac{n-k+q+1}{n+1}\right)^{n+k}\right]$$

$$= \frac{n^m(2m)!}{(2e-5)^m}\left[a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} + \cdots + \frac{a_m}{n^m} + \frac{a_{m+1}}{n^{m+1}} + \cdots\right].$$

In order for $\lim_{n\to\infty} E\left[\left(\frac{n}{c}\right)^m\left(\omega_n - \frac{1}{e}\right)^{2m}\right]$ to exist it is necessary to show that $a_i = 0$, $i = 0, 1, 2, \cdots m-1$. Then $\lim_{n\to\infty} E\left[\left(\frac{n}{c}\right)^m\left(\omega_n - \frac{1}{e}\right)^{2m}\right] = \frac{a_m(2m)!}{(2e-5)^m}$. If we de-

termine the coefficient a_{iq} of n^{-i} in the expansion in powers of n^{-1} of

$$\sum_{k=q+1}^{2m} \frac{(-1)^k n! e^k}{(n+k)!(2m-k)!} \binom{n+1}{k-q} \binom{k-1}{q} \cdot \left(\frac{n-k+q+1}{n+1}\right)^{n+k} = \sum_{i=q}^{\infty} \frac{a_{iq}}{n^i}, \tag{2.2}$$

we will then have

$$a_j = \sum_{q=0}^{j} a_{jq}, \qquad j = 1, 2, \cdots m. \tag{2.3}$$

It can be established at once that $a_0 = 0$ For if we set $q = 0$ in (2 2) and let $n \to \infty$ then (2.2) has the limit $\sum_{k=1}^{2m} \frac{(-1)^k}{(2m-k)!\,k!} = \frac{1}{-(2m)!}$. To determine the expansion of (2.2) in powers of n^{-1} it is sufficient to focus attention on the expansion in powers of n^{-1} of

$$\begin{aligned} &\frac{n!}{(n+k)!}(n+1)(n)\cdots(n-k+q+2)\left(\frac{n-k+q+1}{n+1}\right)^{n+k} \\ &\qquad = \frac{(n+1)(n)\cdots(n-k+q+2)}{(n+k)(n+k-1)\cdots(n+1)}\left(\frac{n-k+q+1}{n+1}\right)^{n+k} \end{aligned}$$

or equivalently on the expansion in powers of x of the function

$$\begin{aligned} &\frac{\left(\frac{1}{x}+1\right)\left(\frac{1}{x}\right)\cdots\left(\frac{1}{x}-k+q+2\right)}{\left(\frac{1}{x}+k\right)\left(\frac{1}{x}+k-1\right)\cdots\left(\frac{1}{x}+1\right)}\left(\frac{\frac{1}{x}-k+q+1}{\frac{1}{x}+1}\right)^{(1/x)+k} \\ &\quad = \frac{x^q(1-x)(1-2x)\cdots(1-(k-q-2)x)}{(1+2x)(1+3x)\cdots(1+kx)}\left(\frac{1-(k-q-1)x}{1+x}\right)^{(1/x)+k} \\ &\quad = x^q(a_{kq0} + a_{kq1}x + a_{kq2}x^2 + \cdots) = x^q F(x). \end{aligned}$$

Here $a_{kq0} = e^{-k+q}$ and the other coefficients may be obtained by a recursion formula. Thus:

$$\begin{aligned} a_{kqp} &= \frac{1}{p!} D_{x=0}^{(p)} F(x) = \frac{1}{p!} D_{x=0}^{(p-1)} [F(x) D \log F(x)] \\ &= \frac{1}{p!} \sum_{s=0}^{p-1} \binom{p-1}{s} D_{x=0}^{(p-s-1)} F(x) D_{x=0}^{(s+1)} \log F(x). \end{aligned}$$

But

$$\begin{aligned} D_{x=0}^{(s+1)} \log F(x) &= D_{x=0}^{(s+1)} \left[\left(\frac{1}{x}+k\right)\log(1-(k-q-1)x)\right. \\ &\quad \left. - \left(\frac{1}{x}+k\right)\log(1+x) + \sum_{i=1}^{k-q-2}\log(1-ix) - \sum_{i=2}^{k}\log(1+ix)\right] \\ &= s!\left[(k-q-1)^{s+1}\left(\frac{k-q-1}{s+2} - 2k+q+1\right)\right. \\ &\quad \left. + (-1)^s\left(1-k-\frac{1}{s+2}\right) - \sum_{i=1}^{k-q-2} i^{s+1} - \sum_{i=2}^{k}(-1)^s i^{s+1}\right] = s!\, b_{kqs}, \end{aligned}$$

so that

$$a_{kqp} = \frac{1}{p!}\sum_{s=0}^{p-1}\binom{p-1}{s}(p-s-1)!\,a_{kq(p-s-1)}\,s!\,b_{kqs} = \frac{1}{p}\sum_{s=0}^{p-1} a_{kq(p-s-1)}b_{kqs}.$$

Of b_{kqs} we need merely notice that it is a polynomial in k of degree $s + 2$ and that $b_{kq0} = -\frac{5}{2}k^2 + Ak + B$, where A and B depend on q only. We wish to determine the value of $a_{kq(i-q)}$ and to this end we solve the system of linear equations

$$\begin{aligned} a_{kq0} \qquad\qquad\qquad\qquad &= e^{-k+q}, \\ \frac{1}{p}\sum_{s=0}^{p-1} a_{kq(p-s-1)}b_{kqs} - a_{kqp} &= 0, \qquad p = 1, 2, \cdots, i - q. \end{aligned}$$

$a_{kq(i-q)}$ is therefore a quotient of two determinants. The determinant in the denominator has the value $(-1)^{i-q}$ while the determinant in the numerator can be expanded by its last column and is therefore the product of $(-1)^{i-q}e^{-k+q}$ and a determinant B_{kqi} whose entries $d_{\alpha\beta}$, $\alpha, \beta = 1, 2, \cdots i - q$, can be described as follows If $\beta > \alpha + 1$ then $d_{\alpha\beta} = 0$. $d_{\alpha(\alpha+1)} = -1$ and when $\beta \leqq \alpha$, $d_{\alpha\beta} = \frac{1}{\alpha}b_{kq(\alpha-\beta)}$, a polynomial of degree $\alpha - \beta + 2$. Thus $a_{kq(i-q)} = e^{-k+q}B_{kqi}$. The determinant B_{kqi} is a polynomial of degree $2(i - q)$ in k and the term of this degree comes only from the product of the diagonal elements. For $B_{kqi} = |d_{\alpha\beta}| = \Sigma \pm \prod_{\alpha=1}^{i-q} d_{\alpha\sigma(\alpha)}$ where $\sigma(\alpha) \leqq \alpha + 1$ and $(\sigma(1), \sigma(2), \cdots \sigma(i - q))$ is a permutation of $(1, 2, \cdots i - q)$. The term $\prod_{\alpha=1}^{i-q} d_{\alpha\sigma(\alpha)}$ has degree $\sum_{\alpha=1}^{i-q}(\alpha - \sigma(\alpha) + \delta(\alpha)) = \sum_{\alpha=1}^{i-q}\delta(\alpha)$ where $\delta(\alpha) = 2$ if $\sigma(\alpha) \leqq \alpha$ and $\delta(\alpha) = 1$ if $\sigma(\alpha) = \alpha + 1$. But $\sum_{\alpha=1}^{i-q}\delta(\alpha) = 2(i - q) \leftrightarrow \delta(\alpha) = 2 \leftrightarrow \sigma(\alpha) \leqq \alpha \leftrightarrow \sigma(\alpha) = \alpha$, so that it is the product of the diagonal terms and only that product which gives to the term of degree $2(i - q)$ in the expansion. Thus

$$\begin{aligned} B_{kqi} &= \frac{1}{(i-q)!}(b_{kq0})^{i-q} + \text{terms of lower degree in } k \\ &= \frac{1}{(i-q)!}\left(-\frac{5}{2}\right)^{i-q} k^{2(i-q)} + \sum_{j=0}^{2(i-q)-1} A_j k^j. \end{aligned}$$

We are now in position to evaluate a_{iq}.

$$\begin{aligned} a_{iq} &= \sum_{k=q+1}^{2m} \frac{(-1)^k e^k}{(2m-k)!(k-q)!}\binom{k-1}{q} a_{kq(i-q)} \\ &= \sum_{k=q+1}^{2m} \frac{(-1)^k e^q}{(2m-k)!(k-q)!}\binom{k-1}{q} B_{kqi} \\ &= \frac{e^q}{(i-q)!}\left(-\frac{5}{2}\right)^{i-q} \sum_{k=q+1}^{2m} \frac{(-1)^k k^{2(i-q)}}{(2m-k)!(k-q)!}\binom{k-1}{q} \\ &\quad + \sum_{k=q+1}^{2m} \frac{(-1)^k e^q}{(2m-k)!(k-q)!}\binom{k-1}{q}\left[\sum_{j=0}^{2(i-q)-1} A_j k^j\right]. \end{aligned} \tag{2.4}$$

To complete the evaluation of a_{iq} we observe that

$$(2.5)\quad \sum_{k=q+1}^{2m} \frac{(-1)^k k^l}{(2m-k)!(k-q)!}\binom{k-1}{q} = \begin{cases} \dfrac{1}{q!} & \text{if } l = 2(m-q), \\ 0 & \text{if } l < 2(m-q). \end{cases}$$

(2.5) implies that $a_{iq} = 0$ if $i < m$ and therefore $a_j = 0$ if $j < m$. The proof of (2.5) is brief. We note that $k^{l-1} = \sum_{j=0}^{l-1} c_j \binom{k+j}{j}$, where c_j is independent of k and $c_{l-1} = (l-1)!$. Then

$$\begin{aligned} \sum_{k=q+1}^{2m} \frac{(-1)^k k^l}{(2m-k)!(k-q)!}\binom{k-1}{q} &= \sum_{k=q+1}^{2m} \frac{(-1)^k k^{l-1} k!}{q!(k-q-1)!}\,\frac{1}{(2m-q)!}\binom{2m-q}{k-q} \\ &= \sum_{j=0}^{l-1}\sum_{k=q+1}^{2m} (-1)^k \frac{c_j k!}{(2m-q)!\,q!\,(k-q-1)!}\binom{2m-q}{k-q}\binom{k+j}{j} \\ &= \sum_{j=0}^{l-1} \frac{c_j (j+q+1)!}{(2m-q)!\,j!\,q!}\left[\sum_{k=q+1}^{2m}(-1)^k\binom{2m-q}{k-q}\binom{k+j}{q+j+1}\right]. \end{aligned}$$

The expression within the brackets is the coefficient of x^{2m-q-1} in $(1-x)^{2m-q}\dfrac{1}{(1-x)^{q+j+2}} = (1-x)^{2m-2q-j-2}$ and this is zero if $j < 2(m-q)-1$ and 1 if $j = 2(m-q)-1$. Accordingly

$$\sum_{k=q+1}^{2m} \frac{(-1)^k k^l}{(2m-q)!(k-q)!}\binom{k-1}{q}$$

$$= \begin{cases} 0 & \text{if } l = 2(m-q), \\ \dfrac{[2(m-q)-1]!\,[2(m-q)-1+q+1]!}{(2m-q)!\,[2(m-q)-1]!\,q!} = \dfrac{1}{q!} & \text{if } l = 2(m-q), \end{cases}$$

and (2.5) is established. Returning to (2.4), $a_{iq} = 0$ when $i < m$, while

$$a_{mq} = \frac{e^q}{(m-q)!\,q!}\left(-\frac{5}{2}\right)^{m-q} = \frac{1}{m!}\binom{m}{q}\left(-\frac{5}{2}\right)^{m-q} e^q;$$

and now applying this expression to (2.3)

$$a_m = \sum_{q=0}^{m} \frac{1}{m!}\binom{m}{q}\left(-\frac{5}{2}\right)^{m-q} e^q = \frac{1}{m!\,2^m}(2e-5)^m.$$

Thus

$$\lim_{n\to\infty} E\left[\left(\frac{n}{c}\right)^m\left(\omega_n - \frac{1}{e}\right)^{2m}\right] = \frac{a_m (2m)!}{(2e-5)!} = \frac{(2m)!}{m!\,2^m},$$

and these are precisely the even order moments of the normal distribution. Thus $\left(\dfrac{n}{c}\right)^{1/2}\left(\omega_n - \dfrac{1}{e}\right)$ is asymptotically normal and so is $\dfrac{\omega_n - E(\omega_n)}{D(\omega_n)}$.

The skewness $\beta_1 = \left(\frac{\mu_3}{\sigma^3}\right)^2$ and kurtosis $\beta_2 = \frac{\mu_4}{\sigma^4}$ of the standardized variable $\frac{\omega_n - E(\omega_n)}{D(\omega_n)}$ are

$$\beta_1 = \frac{1}{n} \frac{(6e^2 - 42e + 70)^2}{(2e - 5)^3} + O(n^{-2}) = \frac{.356 \cdots}{n} + O(n^{-2}),$$

$$\beta_2 = 3 + \frac{1}{n} \frac{24e^3 - 336e^2 + 1368e - 1718}{(2e - 5)^2} + O(n^{-2}) = 3 - \frac{1\ 05 \cdots}{n} + O(n^{-2}).$$

3. Consistency. According to previous discussion in order to prove the consistency of the test for goodness of fit based on the asymptotically normal variable $\frac{\omega_n - E(\omega_n)}{D(\omega_n)}$ it is sufficient to show that, if $x_1, x_2, \cdots, x_n$ is an ordered sample from a population whose distribution function is $G(x)$, then the limiting mean of the random variable $\frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right|$ is not equal to e^{-1} if $F(x) \not\equiv G(x)$ and the limiting variance of this variable is zero This is the content of the next two theorems. In connection with these theorems it is to be observed that, when $y = F(x)$ is continuous, $F^{-1}(y)$, $0 \leqq y \leqq 1$, can be defined unambiguously by writing $F^{-1}(y) = [\text{Sup } x \cdot y = F(x)]$ except for $y = 0$, and $F^{-1}(0) = -\infty$. The function $k(x) = GF^{-1}(x)$ is then a non-decreasing function mapping [0, 1] into [0, 1] and such that $k(0) = 0$ and $k(1) = 1$. Now if $F'(x)$ exists for all but a finite number of points and is never zero then $F^{-1}(x)$ is continuous and so is $k(x)$. If further $G'(x)$ and $F'(x)$ exist and are continuous except for a finite number of points then $(F'(x) \neq 0) k'(x)$ enjoys the same property. These remarks justify the substitutions and partial integrations that are effected in the course of the next two theorems.

THEOREM 3. *Let $F(x)$ and $G(x)$ be continuous distribution functions whose derivatives exist and are continuous except for a finite number of points. If $x_1, x_2, \cdots x_n$ is an ordered sample of n values from the population whose distribution function is $G(x)$ then $(k(x) = GF^{-1}(x))$*

$$E(\Omega_n) = E\left(\frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right|\right)$$
$$= \int_0^{n/(n+1)} \left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n dx \to \int_0^1 e^{-k'(x)}\, dx.$$

The integral $\int_0^1 e^{-k'(x)}\, dx$ has, relative to the class of monotonic functions such that $k(0) = 0$ and $k(1) = 1$, the minimum value e^{-1} and assumes that value only when $k(x) \equiv x$ i.e. $F(x) \equiv G(x)$.

Let us suppose first that $F'(x) \neq 0$ Then $F^{-1}(x)$ is continuous and it is differentiable at all but a finite number of points as is also the function $GF^{-1}(x) = k(x)$.

$$
\begin{aligned}
E(\Omega_n) &= \frac{1}{2}\sum_{i=1}^{n+1} E\left(\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|\right) \\
&= \frac{1}{2}E\left(\left|F(x_1) - \frac{1}{n+1}\right|\right) + \frac{1}{2}E\left(\left|1 - F(x_n) - \frac{1}{n+1}\right|\right) \\
&\qquad + \frac{1}{2}\sum_{i=2}^{n} E\left(\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|\right). \qquad (3.1)
\end{aligned}
$$

The joint probability density element of x_{i-1} and x_i is

$$
\frac{n!}{(i-2)!\,(n-i)!}\, G(x_{i-1})^{i-2}(1 - G(x_i))^{n-i}\, dG(x_{i-1})\, dG(x_i)
$$

in the domain $-\infty < x_{i-1} < x_i < +\infty$ and zero outside that domain. Hence

$$
\begin{aligned}
&\frac{1}{2}\sum_{i=2}^{n} E\left(\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|\right) \\
&\quad = \frac{1}{2}\sum_{i=2}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{x_i}\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right| \\
&\qquad \cdot \frac{n!}{(i-2)!\,(n-i)!}\, G(x_{i-1})^{i-2}(1 - G(x_i))^{n-i}\, dG(x_{i-1})\, dG(x_i) \\
&\quad = \frac{1}{2}n(n-1)\int_{-\infty}^{\infty}\int_{-\infty}^{Y}\left|F(Y) - F(X) - \frac{1}{n+1}\right| \\
&\qquad \cdot [1 - G(Y) + G(X)]^{n-2}\, dG(X)\, dG(Y),
\end{aligned}
$$

and making the transformation $y = F(Y)$ and $x = F(X)$ the integral on the right can be written

$$
\begin{aligned}
&\tfrac{1}{2}n(n-1)\int_0^1\int_0^y\left|y - x - \frac{1}{n+1}\right|[1 - k(y) + k(x)]^{n-2}\, dk(x)\, dk(y) \\
&= \tfrac{1}{2}n(n-1)\int_0^1\int_0^y\left(x - y + \frac{1}{n+1}\right)[1 - k(y) + k(x)]^{n-2}\, dk(x)\, dk(y) \\
&+ n(n-1)\int_{1/n+1}^{1}\int_0^{y-(1/n+1)}\left(y - x - \frac{1}{n+1}\right)[1 - k(y) + k(x)]^{n-2}\, dk(x)\, dk(y).
\end{aligned}
$$

Integrating partially with respect to x, the expression on the right becomes

$$
\begin{aligned}
&\frac{n}{2}\int_0^1\frac{1}{n+1}\, dk(y) - \frac{n}{2}\int_0^1\left(-y + \frac{1}{n+1}\right)[1 - k(y)]^{n-1}\, dk(y) \\
&\qquad - \frac{n}{2}\int_0^1\int_0^y[1 - k(y) + k(x)]^{n-1}\, dx\, dk(y) \\
&\qquad - n\int_{1/n+1}^{1}\left(y - \frac{1}{n+1}\right)[1 - k(y)]^{n-1}\, dk(y) \\
&\qquad\qquad + n\int_{1/n+1}^{1}\int_0^{y-(1/n+1)}[1 - k(y) + k(x)]^{n-1}\, dx\, dk(y),
\end{aligned}
$$

and now integrating with respect to y

$$\frac{1}{2}\sum_{i=2}^{n} E\left(\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|\right) = -\frac{1}{n+1} + \frac{1}{2}\int_0^1 [1 - k(x)]^n \, dx$$
$$+ \frac{1}{2}\int_0^1 k(x)^n \, dx - \int_{1/n+1}^1 [1 - k(x)]^n \, dx - \int_0^{n/n+1} k(x)^n \, dx$$
$$+ \int_0^{n/n+1} \left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n dx.$$

The other two terms in (3.1) are treated similarly. The probability density element of x_1 is $n(1 - G(x_1))^{n-1} \, dG(x_1)$ so that

$$\tfrac{1}{2}E\left(\left|F(x_1) - \frac{1}{n+1}\right|\right) = \frac{n}{2}\int_{-\infty}^{\infty}\left|F(x) - \frac{1}{n+1}\right| (1 - G(x))^{n-1} \, dG(x)$$
$$= \frac{n}{2}\int_0^1 \left|x - \frac{1}{n+1}\right| (1 - k(x))^{n-1} \, dk(x)$$
$$= \frac{1}{2(n+1)} - \frac{1}{2}\int_0^{1/n+1} (1 - k(x))^n \, dx$$
$$+ \frac{1}{2}\int_{1/n+1}^1 (1 - k(x))^n \, dx.$$

Similarly we find that

$$\frac{1}{2} E\left(\left|1 - F(x_n) - \frac{1}{n+1}\right|\right) = \frac{1}{2(n+1)}$$
$$+ \frac{1}{2}\int_0^{n/n+1} k(x)^n \, dx - \frac{1}{2}\int_{1/n+1}^1 k(x)^n \, dx.$$

Thus

$$E(\Omega_n) = \int_0^{n/n+1} \left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n dx.$$

This result is, however, independent of the hypothesis $F'(x) \neq 0$. For if $F'(x)$ is sometimes zero we may select a sequence of distribution functions $F_m(x)$, $m = 1, 2, \cdots$, which converges everywhere to $F(x)$ and which is such that $F_m'(x) \neq 0$. The $F_m(x)$ otherwise satisfy the conditions of the theorem. If Ω_{mn} is that function of $x_1, x_2, \cdots, x_n$ obtained by replacing $F(x)$ by $F_m(x)$ in Ω_n then Ω_{mn} converges to Ω_n for every fixed set of $x_1, x_2, \cdots, x_n$ and $E(\Omega_{mn})$ converges to $E(\Omega_n)$ since both Ω_{mn} and Ω_n are bounded by 1. Furthermore if x_0 is any value such that $F'(x_0) \neq 0$ and $y_0 = F(x_0)$ then $F_m^{-1}(y_0)$ converges to $F^{-1}(y_0) = x_0$. For if x_1 is a cluster point of the set $F_m^{-1}(y_0)$, then there exists, for a given ϵ, a sufficiently large m such that $|F(x_1) - F_m(x_1)| < \epsilon$ (because $F_m(x) \to F(x)$) while, for the same m, $|F_m(x_1) - y_0| < \epsilon$ because of the continuity of $F_m(x)$ Thus $|F(x_1) - y_0| < 2\epsilon$ and, since ϵ is arbitrary, $y_0 = F(x_1) = F(x_0)$. So $x_1 = x_0$ since $F'(x_0) \neq 0$ Thus $F_m^{-1}(y) \to F^{-1}(y)$ for any

value of y such that if x is mapped into y by $F(x)$ then $F'(x) \neq 0$. This set on the y axis however includes all y except for a set of measure zero and so $F_m^{-1}(y) \to F^{-1}(y)$ almost everywhere. So $k_m(y) = GF_m^{-1}(y) \to GF^{-1}(y) = k(y)$ almost everywhere and

$$\left[1 - k_m\left(y + \frac{1}{n+1}\right) + k_m(y)\right]^n \to \left[1 - k\left(y + \frac{1}{n+1}\right) + k(y)\right]^n$$

almost everywhere. Then

$$\int_0^{n/n+1} \left[1 - k_m\left(x + \frac{1}{n+1}\right) + k_m(x)\right]^n dx$$
$$\to \int_0^{n/n+1} \left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n dx$$

since both integrands are bounded by 1. Therefore the equality

$$E(\Omega_{mn}) = \int_0^{n/n+1} \left[1 - k_m\left(x + \frac{1}{n+1}\right) + k_m(x)\right]^n dx$$

is preserved as $m \to \infty$.

Now $k(x)$ is a monotonic function and hence has a derivative almost everywhere. Then

$$\left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n$$
$$= \left[1 - \frac{1}{n+1}\left(k\left(x + \frac{1}{n+1}\right) - k(x) \Big/ \frac{1}{n+1}\right)\right]^n$$

converges to $e^{-k'(x)}$ almost everywhere. If we write

$$H_n(x) = \left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n$$

when $0 \leqq x \leqq \dfrac{n}{n+1}$ and $H_n(x) = 0$ when $\dfrac{n}{n+1} < x \leqq 1$, then

$$\int_0^1 H_n(x)\,dx = \int_0^{n/n+1} \left[1 - k\left(x + \frac{1}{n+1}\right) + k(x)\right]^n dx \to \int_0^1 e^{-k'(x)}\,dx$$

as $n \to \infty$. The curve $y = e^{-x}$ lies always above its tangents and the tangent at $x = 1$ is $y = -\dfrac{1}{e}x + \dfrac{2}{e}$. Thus $e^{-x} \geqq -\dfrac{1}{e}x + \dfrac{2}{e}$ for all x, equality holding only when $x = 1$, and therefore $e^{-k'(x)} \geqq -\dfrac{1}{e}k'(x) + \dfrac{2}{e}$, equality holding only when $k'(x) = 1$.

So

$$\int_0^1 e^{-k'(x)}\,dx \geqq -\frac{1}{e}\int_0^1 k'(x)\,dx + \frac{2}{e},$$

equality holding if and only if $k'(x) = 1$ almost everywhere. But for any monotonic non-decreasing function

$$\int_0^1 k'(x)\,dx \leqq k(1) - k(0),$$

equality holding if and only if $k(x)$ is absolutely continuous. Hence

$$\int_0^1 e^{-k'(x)}\,dx \geqq -\frac{1}{e}\int_0^1 k'(x)\,dx + \frac{2}{e} \geqq \frac{1}{e},$$

and the equality runs through if and only if $k(x)$ is an absolutely continuous function such that $k'(x) = 1$ almost everywhere But this is true of $k(x)$ if and only if $k(x) \equiv x$ and this in turn is true if and only if $F(x) \equiv G(x)$

THEOREM 4. *The random variable Ω_n has limiting variance zero; i.e.,* $\lim_{n\to\infty} E(\Omega_n^2) = \left[\int_0^1 e^{-k'(x)}\,dx\right]^2$.

As before we assume first that $F'(x) \not\equiv 0$. Then

$$E(\Omega_n^2) = E\left[\left(\frac{1}{2}\sum_{i=2}^{n}\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|\right)^2\right]$$

$$\text{(4.1)} \qquad + E\left[\left|F(x_1) - \frac{1}{n+1}\right|\Omega_n\right] + E\left[\left|1 - F(x_n) - \frac{1}{n+1}\right|\Omega_n\right]$$

$$- E\left[\frac{1}{4}\left(\left|F(x_1) - \frac{1}{n+1}\right| + \left|1 - F(x_n) - \frac{1}{n+1}\right|\right)^2\right].$$

Suppose [Sup x: $k(x) = 0$] $= a$ and [Inf x: $k(x) = 1$] $= b$. We may then obtain $\lim_{n\to\infty} E\left[\left|F(x_1) - \frac{1}{n+1}\right|\Omega_n\right]$ in the following manner:

$$\left|E\left[\left|F(x_1) - \frac{1}{n+1}\right|\Omega_n\right] - E[a\Omega_n]\right| \leqq E\left[\left|\left|F(x_1) - \frac{1}{n+1}\right| - a\right|\Omega_n\right]$$

$$\text{(4.2)} \qquad \leqq E\left[\left|F(x_1) - \frac{1}{n+1} - a\right|\Omega_n\right]$$

$$\leqq \left[E\left(F(x_1) - \frac{1}{n+1} - a\right)^2\right]^{1/2}[E(\Omega_n^2)]^{1/2}.$$

But $\Omega_n \leqq 1$ so that $E(\Omega_n^2)$ is bounded as $n \to \infty$. On the other hand

$$E\left(F(x_1) - \frac{1}{n+1} - a\right)^2 = n\int_{-\infty}^{\infty}\left(F(x_1) - \frac{1}{n+1} - a\right)^2 (1 - G(x_1))^{n-1}\,dG(x_1)$$

$$= n\int_0^1 \left(x - a - \frac{1}{n+1}\right)^2 (1 - k(x))^{n-1}\,dk(x)$$

$$= \left(a + \frac{1}{n+1}\right)^2 + \int_0^1 2\left(x - a - \frac{1}{n+1}\right)(1 - k(x))^n\,dx.$$

As $n \to \infty$ the expression on the right tends to $a^2 + \int_0^a 2(x - a)\,dx = 0$. Thus the expression on the right of (4.2) goes to zero as $n \to \infty$ and therefore

$$\lim_{n\to\infty} E\left[\left|F(x_1) - \frac{1}{n+1}\right|\Omega_n\right] = \lim_{n\to\infty} E\,[a\Omega_n] = a\int_0^1 e^{-k'(x)}\,dx. \tag{4.3}$$

In a similar manner we obtain

$$\lim_{n\to\infty} E\left[\left|1 - F(x_n) - \frac{1}{n+1}\right|\Omega_n\right] = (1 - b)\int_0^1 e^{-k'(x)}\,dx \tag{4.4}$$

and

$$\lim_{n\to\infty} - E\left[\frac{1}{4}\left(\left|F(x_1) - \frac{1}{n+1}\right| + \left|1 - F(x_n) - \frac{1}{n+1}\right|\right)^2\right] = -\tfrac{1}{4}(a + 1 - b)^2 \tag{4.5}$$

The first term on the right of (4.1) remains to be investigated. We have

$$\begin{aligned} E\left[\left(\frac{1}{2}\sum_{i=2}^{n}\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|\right)^2\right] &= \frac{1}{4}E\left[\sum_{i=2}^{n}\left(F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right)^2\right] \\ &+ \frac{1}{2}E\left[\sum_{i=2}^{n-2}\sum_{j=i+2}^{n}\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|\left|F(x_j) - F(x_{j-1}) - \frac{1}{n+1}\right|\right] \\ &+ \frac{1}{2}E\left[\sum_{i=2}^{n-1}\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|\left|F(x_{i+1}) - F(x_i) - \frac{1}{n+1}\right|\right]. \end{aligned} \tag{4.6}$$

The joint probability density element of x_{i-1} and x_i is

$$\frac{n!}{(i-2)!\,(n-i)!}(1 - G(x_{i-1}))^{i-2}G(x_i)^{n-i}\,dG(x_{i-1})\,dG(x_i)$$

so that

$$\begin{aligned} &\frac{1}{4}E\left[\sum_{i=2}^{n}\left(F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right)^2\right] \\ &= \frac{1}{4}n(n-1)\iint_{-\infty<X<Y<\infty}\left(F(Y) - F(X) - \frac{1}{n+1}\right)^2 \\ &\quad\cdot [1 - G(Y) + G(X)]^{n-2}\,dG(X)\,dG(Y) \\ &= \frac{1}{4}n(n-1)\int_0^1\int_0^y\left(y - x - \frac{1}{n+1}\right)^2[1 - k(y) + k(x)]^{n-2}\,dk(x)\,dk(y). \end{aligned}$$

In this latter double integral we integrate first with respect to x and then with respect to y obtaining

$$\frac{-n-3}{4(n+1)^2} - \frac{1}{2}\int_0^1 \left(y - \frac{1}{n+1}\right)[1 - k(y)]^n\, dy - \frac{1}{2}\int_0^1 \left(\frac{n}{n+1} - x\right) k(x)^n\, dx + \frac{1}{2}\iint_{0<x<y<1} [1 - k(y) + k(x)]^n\, dx\, dy,$$

and proceeding to the limit

$$\begin{aligned}(4.7)\quad &\lim_{n\to\infty} \frac{1}{4} E\left[\sum_{i=2}^{n}\left(F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right)^2\right] \\ &= -\frac{1}{2}\int_0^a y\, dy - \frac{1}{2}\int_b^1 (1 - x)\, dx + \frac{1}{2}\iint_{\substack{0<x<y<1\\ k(x)=k(y)}} dx\, dy \\ &= -\frac{1}{4}a^2 - \frac{1}{4}(1 - b)^2 + \frac{1}{2}\iint_{\substack{0<x<y<1\\ k(x)=k(y)}} dx\, dy.\end{aligned}$$

The joint probability density element of x_{i-1}, x_i, x_{j-1}, x_j when $j > i + 1$ is

$$\frac{n!}{(i-2)!(j-i-2)!(n-j)!} G(x_{i-1})^{i-2}[G(x_{j-1}) - G(x_i)]^{j-i-2} [1 - G(x_j)]^{n-j}\, dG(x_{i-1})\, dG(x_i)\, dG(x_{j-1})\, dG(x_j),$$

so

$$\begin{aligned}&\frac{1}{2} E\left[\sum_{i=2}^{n-2}\sum_{j=i+2}^{n}\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|\left|F(x_j) - F(x_{j-1}) - \frac{1}{n+1}\right|\right] \\ &= \frac{1}{2}n(n-1)(n-2)(n-3)\iiiint_{0<X<Y<U<V<1}\left|F(Y) - F(X) - \frac{1}{n+1}\right| \\ (4.8)\quad &\cdot\left|F(V) - F(U) - \frac{1}{n+1}\right|[1 - G(V) + G(U) \\ &- G(Y) + G(X)]^{n-4}\, dG(X)\, dG(Y)\, dG(U)\, dG(V) \\ &= \frac{1}{2}n(n-1)(n-2)(n-3)\iiiint_{0<x<y<u<v<1}\left|y - x - \frac{1}{n+1}\right|\left|v - u - \frac{1}{n+1}\right| \\ &\cdot [1 - k(v) + k(u) - k(y) + k(x)]^{n-4}\, dk(x)\, dk(y)\, dk(u)\, dk(v).\end{aligned}$$

The joint probability density element of x_{i-1}, x_i, x_{i+1} is

$$\frac{n!}{(i-2)!(n-i-1)!} G(x_{i-1})^{i-2}[1 - G(x_{i+1})]^{n-i-1}\, dG(x_{i-1})\, dG(x_i)\, dG(x_{i+1})$$

and so

$$\frac{1}{2} E\left[\sum_{i=2}^{n-1}\left|F(x_i) - F(x_{i-1}) - \frac{1}{n+1}\right|\left|F(x_{i+1}) - F(x_i) - \frac{1}{n+1}\right|\right]$$

$$= \frac{1}{2} n(n-1)(n-2) \iiint\limits_{0<X<Y<V<1} \left|F(Y) - F(X) - \frac{1}{n+1}\right|$$

$$\left|F(V) - F(Y) - \frac{1}{n+1}\right| [1 - G(V) + G(X)]^{n-3}\, dG(X)\, dG(Y)\, dG(V) \tag{4.9}$$

$$= \frac{1}{2} n(n-1)(n-2) \iiint\limits_{0<x<y<v<1} \left|y - x - \frac{1}{n+1}\right|$$

$$\cdot \left|v - y - \frac{1}{n+1}\right| [1 - k(v) + k(x)]^{n-3}\, dk(x)\, dk(y)\, dk(v)$$

We introduce the symbol $S(p, q)$ as follows

$$S(p, q) = \begin{cases} 1 & \text{if } q \leqq p + \dfrac{1}{n+1}, \\ -1 & \text{if } q > p + \dfrac{1}{n+1}. \end{cases}$$

Then in the integral on the right of (4.8) we perform a partial integration with respect to u and add to the integral on the right of (4.9) We get

$$\frac{1}{2} n(n-1)(n-2) \iiint\limits_{0<x<y<v<1} \frac{1}{n+1}\left|y - x - \frac{1}{n+1}\right|$$

$$\cdot [1 - k(y) + k(x)]^{n-3}\, dk(x)\, dk(y)\, dk(v)$$

$$- \frac{1}{2} n(n-1)(n-2) \iiiint\limits_{0<x<y<u<v<1} S(u, v)\left|y - x - \frac{1}{n+1}\right|$$

$$\cdot [1 - k(v) + k(u) - k(y) + k(x)]^{n-3}\, dk(x)\, dk(y)\, dk(v)\, du,$$

and now integrating with respect to v in the triple integral and performing partial integrations with respect to x and collecting terms the sum of (4.8) and (4.9) becomes

$$\frac{n(n-1)}{4(n+1)^2} - \frac{n(n-1)}{2(n+1)} \int_0^1 \left|y - \frac{1}{n+1}\right| [1 - k(y)]^{n-1}\, dk(y) - \frac{2n(n-1)}{n+1}$$

$$\cdot \iint\limits_{0<x<y<1} S(x, y)[1 - k(y) + k(x)]^{n-1}\, dx\, dk(y) + \tfrac{1}{2}n(n-1)$$

$$\cdot \iiint\limits_{0<y<u<v<1} S(u, v)\left|y - \frac{1}{n+1}\right|[1 - k(v) + k(u) - k(y)]^{n-2}$$

$$\cdot\, dk(y)\, dk(v)\, du + \tfrac{1}{2}n(n-1)$$

$$\cdot \iiiint\limits_{0<x<y<u<v<1} S(u, v)S(x, y)[1 - k(v) + k(u) - k(y) + k(x)]^{n-2}\, dk(y)\, dk(v)\, dx\, du.$$

Now some tedious, although in principle straightforward, calculations show that the first three terms of this expression approach

$$-\tfrac{1}{4} - \tfrac{1}{2}a - \tfrac{1}{2}(1-b) + \int_0^1 e^{-k'(x)}\,dx, \tag{4.10}$$

that the triple integral approaches

$$\tfrac{1}{2}a + \tfrac{1}{2}a(1-b) + \tfrac{1}{2}a^2 - a\int_0^1 e^{-k'(u)}\,du, \tag{4.11}$$

and that the quadruple integral approaches

$$\begin{aligned} 2\iint\limits_{0<x<u<1} e^{-k'(x)-k'(u)}\,dx\,du - \int_0^1 e^{-k'(x)}\,dx - (1-b)\int_0^1 e^{-k'(x)}\,dx \\ - \frac{1}{2}\iint\limits_{\substack{0<x<u<1\\ k(x)=k(u)}} dx\,du + (1-b)^2 + \tfrac{1}{2}b(1-b) + \tfrac{1}{4}. \end{aligned} \tag{4.12}$$

Thus collecting the results of (4.3), (4.4), (4.5), (4.7), (4.10), (4.11), and (4.12) we have

$$\lim_{n\to\infty} E(\Omega_n^2) = 2\iint\limits_{0<x<u<1} e^{-k'(x)-k'(u)}\,dx\,du.$$

Since the integrand is symmetrical in the variables u and x we may write

$$\lim_{n\to\infty} E(\Omega_n^2) = \iint\limits_{\substack{0<x<1\\ 0<u<1}} e^{-k'(x)-k'(u)}\,dx\,du = \left[\int_0^1 e^{-k'(x)}\,dx\right]^2, \tag{4.13}$$

and this proves the theorem in the case $F'(x) \neq 0$

Using the procedure of theorem 3 we may however extend the theorem to include the possibility that $F'(x)$ is sometimes zero. But it must be shown additionally that the sequence $F_m(x)$ can be so chosen that Ω_{mn} converges to Ω_n uniformly in n, i.e. that, for a given ϵ, $|\Omega_{mn} - \Omega_n| < \epsilon$ for m sufficiently large and for any value of n If this is true then, observing that $0 \leqq \Omega_{mn} + \Omega_n \leqq 2$, $|\Omega_{mn}^2 - \Omega_n^2| < 2\epsilon$ and

$$|E(\Omega_{mn}^2) - E(\Omega_n^2)| \leqq E(|\Omega_{mn}^2 - \Omega_n^2|) \leqq 2\epsilon$$

independently of n. Letting $n \to \infty$

$$\left| \left[\int_0^1 e^{-k_m'(x)}\,dx\right]^2 - \lim_{n\to\infty} E(\Omega_n^2) \right| \leqq 2\epsilon,$$

and now letting $m \to \infty$ (the $F_m(x)$ constructed below are such that $k_m'(x) \to k'(x)$)

$$\left| \left[\int_0^1 e^{-k'(x)}\,dx\right]^2 - \lim_{n\to\infty} E(\Omega_n^2) \right| \leqq 2\epsilon.$$

Since ϵ is arbitrary this implies (4.13), so that the theorem is extended to include the possibility that $F'(x)$ is sometimes zero. That the sequence $F_m(x)$ can be chosen so that Ω_{mn} converges to Ω_n uniformly in n can be shown as follows. The set of points on the x axis for which $F'(x) = 0$ maps into a set of points on the y axis of measure zero For any m we may enclose this set on the y axis in an open set S of measure less than $\frac{1}{m}$ S is the union of disjoint open intervals S_i, $i = 1, 2, \cdots$. The sets $T_i = F^{-1}(S_i)$ on the x axis are disjoint open intervals Now we may construct a distribution function $F_m(x)$ which coincides with $F(x)$ outside ΣT_i, is such that $F'_m(x) \neq 0$, and otherwise satisfies the conditions of the theorem (stated explicitly in Theorem 3). The sequence $F_m(x)$ converges to $F(x)$. Furthermore

$$\begin{aligned} | \Omega_{mn} - \Omega_n | & \\ &= \left| \frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| - \frac{1}{2} \sum_{i=1}^{n+1} \left| F_m(x_i) - F_m(x_{i-1}) - \frac{1}{n+1} \right| \right. \\ &\leqq \frac{1}{2} \sum_{i=1}^{n+1} \left| \left| F(x_i) - F(x_{i-1}) - \frac{1}{n+1} \right| - \left| F_m(x_i) - F_m(x_{i-1}) - \frac{1}{n+1} \right| \right| \qquad (4.14) \\ &\leqq \frac{1}{2} \sum_{i=1}^{n+1} | [F(x_i) - F(x_{i-1})] - [F_m(x_i) - F_m(x_{i-1})] |. \end{aligned}$$

For any particular set of values of $x_1, x_2, \cdots x_n$ some (possibly none or possibly all) of the x_i will fall into intervals of the ΣT_i. If this finite set of intervals, each containing at least one x_i, is say $T_1, T_2, \cdots, T_k$, then a simple analysis of the sum on the right of (4.14) shows that it is less than twice the total length of the intervals $F(T_1), F(T_2), \cdots F(T_k)$ and this total length is less than $\frac{1}{m}$.

Thus $| \Omega_{mn} - \Omega_n | < \frac{1}{m}$ and this result is independent of n

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946

[2] B F Kimball, "Some basic theorems for developing tests of fit for the case of the non-parametric probability distribution function I," *Annals of Math Stat*, Vol. 18 (1947), pp 540–548.

[3] P. A. P Moran, "The random division of an interval," *Jour. Royal Stat. Soc., Supp*, Vol. 9 (1947), pp. 92–98.

[4] N. Smirnoff, "Sur la distribution de ω^2," *Compte Rendus de l'Academie des Sciences*, Paris, 202 (1932), p 449

[5] A. Wald and J Wolfowitz, "On a test of whether two samples are from the same population," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 147–162.

[6] S S. Wilks, *Mathematical Statistics*, Princeton University Press, 1943.

ON A PROBLEM IN THE THEORY OF k POPULATIONS[1]

BY RAGHU RAJ BAHADUR

University of North Carolina

1. Summary. In two recent papers, Paulson [1] and Mosteller [2] have called attention to several unsolved problems in k-sample theory. A problem which is typical of the ones considered in this paper is as follows.

Let $\pi_1, \pi_2, \cdots, \pi_k$ be a set of normal populations, π_i having an unknown mean m_i and variance σ^2, $G(x, \theta_i)$ being the distribution function which characterizes π_i. Samples of equal size are drawn from each population, $\bar{X}_i$ being the sample means, and S^2 the estimate of σ^2 obtained. The problem is to construct a suitable decision rule $d = d(\{\bar{X}_i\}; S^2)$ to select one or more populations, the object being to minimize the expected value of the random distribution function

$$G(x \mid s(d)) = \sum_{i=1}^{k} Z_i(d) \cdot G(x, \theta_i) \Big/ \sum_{i=1}^{k} Z_i(d),$$

where $Z_i(d) = 1$ if π_i is selected by d, and $= 0$ otherwise. It is shown that under the restriction of impartial decision, the rule $d_k =$ "Always select only the population corresponding to the greatest $\bar{X}_i$" cannot be improved, no matter what x or the true parameter values may be. It follows (i) that d_k is the uniformly best decision rule in the class of impartial decision rules for all weight functions of type

$$W = \max_i \{m_i\} - \left(\sum_{i=1}^{k} z_i m_i \Big/ \sum_{i=1}^{k} z_i\right),$$

and (ii) that the customary F and t tests of analysis of variance are not relevant to the problem.

This result is an application of Theorem 1 which applies to a number of similar problems concerning k populations, especially when the populations admit sufficient statistics for their parameters. Two examples of statistical applications are given in Section 6.

2. Introduction. It has been recognized for some time that the classical theory of statistical inference does not provide direct answers to many problems which are of great interest in the applications. One of them, which arises in the theory of samples from k populations, is what Mosteller has called "the problem of the greatest one." The word "population" is used here for a process, $\pi(\theta)$ say, which generates independent random variables $X_1, X_2, \cdots$, each X having the same distribution function $P(X \leq x) = G(x, \theta)$ say, and a set of X's

[1] This paper is based on a thesis submitted to the Department of Mathematical Statistics, University of North Carolina, in partial fulfilment of the requirements for the Ph D. degree This work was sponsored by the Office of Naval Research.

which have been generated by π is called a sample from the population We shall describe the problem, as also the formulation adopted in the following section, in terms of two special cases. These cases occur when the k given populations $\pi_1, \pi_2, \cdots, \pi_k$ are such that π_i is characterized by the distribution function $G(x, \theta_i) = h\left(\frac{x - b_i}{c_i}\right)$, $\theta_i = (b_i, c_i)$, $c_i > 0$, $i = 1, 2, \cdots, k$, where $h(x)$ is an absolutely continuous non-decreasing function with $h(-\infty) = 0$, $h(+\infty) = 1$. Such sets of populations appear frequently in statistical theory and practice, a given set of normal, or rectangular, or gamma type populations being familiar instances.

CASE 1. Let X_{ij}, $j = 1, 2, \cdots, n$ be a sample from the population π_i, $i = 1, 2, \cdots, k$ where π_i is characterized by the distribution function $h\left(\frac{x - b_i}{c}\right)$, b_i being unknown, and suppose that the statistician is asked to select the population which he thinks has the greatest b_i, but is allowed to select more than one population if (as a consequence, say, of "insignificant" outcomes of tests of differences between populations) he does not feel confident enough to select only one This situation will occur if, for example, the X_{ij}'s are observed yields in an agricultural experiment in which each of k varieties has been replaced n times, the yield with variety π_i being normally distributed with unknown mean m_i and variance σ^2, and the statistician is asked to recommend one or more varieties for general use. (Cf. Example 1 in Section 6.)

CASE 2. Suppose now that the X_{ij}'s are samples from populations π_i characterized by distribution functions $h\left(\frac{x - b}{c_i}\right)$, $c_i > 0$ unknown, $i = 1, 2, \cdots, k$, and the statistician is asked to select the population which he thinks has the greatest $1/c_i$, but is allowed to select more than one population.[2] This situation will occur if, for instance, the π_i are factories producing an article having a numerical quality characteristic X, $h\left(\frac{x - b}{c_i}\right)$ being the distribution function of X in the product of π_i, and the statistician is required to assign production to one or more factories, the object being to obtain product of stable quality, b being the standard characteristic

It is clear that the usual statistical theory, which confines itself to estimation of parameters θ_i and testing of hypotheses of the kind $H_0(b_i = \text{constant})$, is inadequate to deal with problems of this sort, where a definite course of action is required of the statistician. It is hardly necessary to add that selection is an important problem in the applications, and the testing of hypotheses is often an indirect attempt to justify selection. In accordance with Wald's formulation of

[2] There is no essential difference between the problem of the greatest one and the problem of the least one. In order to avoid trivial complications, the terminology of the former will be used wherever possible.

the problem of statistical inference,[3] we proceed to consider explicitly the purpose of selection and the "loss" involved in making any particular selection.

3. A class of weight functions. Let $\pi_1, \pi_2, \cdots, \pi_k$ be a given set of populations, π_i being characterized by the distribution function $G(x, \theta_i)$, and let us denote any particular selection, say s, by indicator variables $z_1, z_2, \cdots, z_k$ where $z_i = 1$ if π_i is selected and $= 0$ otherwise Since any meaningful selection must concern itself with the random variables generated by the populations selected, consider the function $G(x \mid s) = \sum_{i=1}^k z_i G(x, \theta_i) / \sum_{i=1}^k z_i$ $G(x \mid s)$ is a distribution function, and provides a logical and direct overall picture of the effect of making the selection s, since no distinction is made between the populations selected. In immediate generalization, we define a "selection" s to be a vector, $s = (p_1, p_2, \cdots, p_k)$ with $p_i \geq 0$, $\sum_{i=1}^k p_i = 1$, and put $G(x \mid s) = \sum_{i=1}^k p_i G(x, \theta_i)$. Roughly speaking, $G(x \mid s)$ is the distribution function which characterizes the mixed population obtained if sampling rates $p_1, p_2, \cdots, p_k$ are assigned to $\pi_1, \pi_2, \cdots, \pi_k$ respectively, $p_r = 0$ corresponding to rejection of π_r. Henceforth, a selection vector will be called a decision.

Now, if each of the $G(x, \theta_i)$'s were known, an appropriate decision s could be chosen without resort to sampling. If not, the statistician must construct (in advance) and use an s-valued function of the sample values. Such a function, say d, is called a statistical decision function or decision rule The decision s according to d, say $s(d) = (p_1(d), p_2(d), \cdots, p_k(d))$, is in general a random vector, so that for any fixed x, $G(x \mid s(d))$ is a random variable. Consider the distribution function $H(x \mid d) = E[G(x \mid s(d))] = \sum_{i=1}^k G(x, \theta_i) E[p_i(d)]$, where E denotes the expectation operator. It represents the average overall effect of using the decision rule d, and so affords a reasonable description of the performance of d Clearly, the problem is to construct d in such a way that $H(x \mid d)$ has desirable properties

The "desirable properties" will depend, of course, on the particular problem being considered Returning to our two cases, denote the arbitrary but given set of all possible parameter points $\omega = (\theta_1, \theta_2, \cdots, \theta_k)$ by Ω, and let D be a given class of decision rules $d = d(\{X_{ij}\})$. Then, in Case 1 we wish to choose $d^* \in D$ such that $H(x \mid d^*) = \inf_{d \in D} H(x \mid d)$ for every x and every $\omega \in \Omega$. In Case 2, we wish to choose d^* so that for every x and every ω we have $H(x \mid d^*) = \inf_{d \in D} H(x \mid d)$ whenever $x < b$, and $= \sup_{d \in D} H(x \mid d)$ whenever $x > b$

These requirements are very strong, and in general no such d^* will exist without heavy restrictions on Ω and on D. (Cf. however the corollary to Theorem 1. It will be found that in a number of cases no restrictions on Ω are required provided that D is the class defined there.) For some purposes, it may be sufficient to consider functionals of $H(x \mid d)$. The functionals which are most useful in the applications are the moments. Thus, one may wish to find d^* such that $\alpha(d^*) = \sup_{d \in D} \alpha(d)$, where $\alpha(d) = \int_{-\infty}^{+\infty} g(x) dH(x \mid d)$, $g(x)$ being some appropriate function.

[3] See, for example, [3], Chapter VI.

For example, in Case 1 we may take $g(x) = x$. Then $\alpha(d)$ is the mean of a random variable having $H(x \mid d)$ for its distribution function, and constructing a suitable d to maximize $\alpha(d)$ is "the problem of the greatest mean." Again, in Case 2 we may take $g(x) = -(x - b)^2$, and in that case maximizing $\alpha(d)$ would be "the problem of the smallest variance."[4]

In terms of mixtures of distributions, $H(x \mid d)$ is the mixture of $G(x \mid s)$ with respect to δ, where δ is the probability measure induced by the decision rule d on the class of Borel sets in the space of all possible decisions s. It follows by the use of Theorem 5 in [4], or otherwise directly, that maximizing $\alpha(d)$ is equivalent to maximizing the expected value (δ) of $\sum_{i=1}^{k} p_i \int_{-\infty}^{+\infty} g(x)dG(x, \theta_i)$. Writing $g_i = \int_{-\infty}^{+\infty} g(x)dG(x, \theta_i)$, one may say that the object is to construct d in such a way that the expected value (δ) of the "weight function"

$$W(\omega, s) = \max_i \{g_i\} - \sum_{i=1}^{k} p_i g_i$$

is minimized for every ω. W represents the "loss" incurred by choosing the decision s when the true parameter point is ω. It will be seen that W defined according to (A) in Section 5 includes essentially all weight functions which are likely to be of interest in the type of problem considered in this paper.

We have so far not emphasized the obvious fact that the probability measure δ which is induced by d on the space of decisions will in general depend on the unknown parameter point ω. Therefore, the expected value (δ) of W is to be written as $E[W(\omega, s(d)) \mid \omega] = r(d \mid \omega)$ say. Following the usual terminology, we shall call $r(d \mid \omega)$ the risk function of the rule d, and shall say that $d^* \in D$ is the uniformly best rule in the class D if $r(d^* \mid \omega) = \inf_{d \in D} r(d \mid \omega)$ for all $\omega \in \Omega$.

4. A class of decision rules. The class of decision rules to which we shall confine ourself is rather limited, and may be described as follows, with reference to the previous sections:

(i) Given independent random variables $\{X_{ij}\}, j = 1, 2, \cdots, n; i = 1, 2, \cdots, k$ from the k populations π_i, let

$$X_i = \phi(X_{i1}, X_{i2}, \cdots, X_{in}), \; i = 1, 2, \cdots, k \text{ and } Y = \psi(\{X_{ij}\}),$$

where $X_1, X_2, \cdots, X_k$; Y is an independent set, and the X_i's have frequency functions. The choice of ϕ and ψ will depend upon particular cases: in Case 1, $X_1, \cdots, X_k$; Y will be statistics relevant to the estimation of

[4] An unpublished theorem of Herbert Robbins insures that if a d^* satisfies the strong requirements of the preceding paragraph, it will also maximize all functionals $\alpha(d)$ corresponding to such functions $g(x)$.

$b_1, b_2, \cdots, b_k; c$ respectively, and in Case 2 they will be relevant to $c_1, c_2, \cdots, c_k, b$.[5]

(ii) Given the statistics $\{X_i\}$; Y, $D(\phi, \psi)$ is the class of all impartial decision rules which are based on them. A decision rule $d = d(\{X_i\}; Y)$ is said to be impartial if it has the following structure. Let $X_{(1)} < X_{(2)} < \cdots < X_{(k)}$ be the ordered X_i's Then d defines non-negative random variables $\lambda_j(X_{(1)}, X_{(2)}, \cdot, X_{(k)}; Y), j = 1, 2, \cdot, k$ such that $\sum_{j=1}^k \lambda_j \equiv 1$, and λ_j is the proportion $p(d)$ which is assigned by d to the π corresponding to $X_{(j)}$. We use the term "impartial" for such decision rules because they determine the proportions $[\lambda_1, \lambda_2, \cdots, \lambda_k]$ without regard to which X belongs to which population, and then assign these proportions in strict order of the X_i's

We shall specify the intuitively plausible class of impartial decision rules for the important normal cases, and give a few instances of such rules.

Suppose first that the X_{ij}'s are from normal populations having means m_i and a common variance σ^2, and that we are interested in the problem of the greatest mean. D is then the class of all impartial decision rules which are based on the statistics

$$X_i = \bar{X}_i = \sum_{j=1}^{n} X_{ij}/n, \qquad i = 1, 2, \cdots, k,$$
$$Y = S^2 = \sum_{j=1}^{k} \sum_{j=1}^{n} (X_{ij} - \bar{X}_i)^2/k\ (n - 1).$$

The numerical factors are of no importance, and may be omitted (Cf. footnote 4. See also Example 2 in Section 6, where such factors have been omitted for convenience). A rather simple member of D is the rule $[\lambda_{k-1} \equiv 1/3, \lambda_k \equiv 2/3]$ i.e. "Always assign the proportion 2/3 to the population which has the greatest $\bar{X}_i$, and the proportion 1/3 to the population with the second greatest." In using this rule although the λ_j's remain constant from sample to sample, the decision $s(d)$ is a random vector In general, however, the λ_j's will themselves be random variables. This is the case if, for instance, one insists on utilising the standard test of differences between populations, and uses the impartial rule "Perform the F test of $H_0(m_i =$ constant) at the five per cent level. If H_0 is rejected, assign the proportion 1 to the population which has the greatest $\bar{X}_i$. If not, assign equal proportions to all populations for which $\bar{X}_i > \sum_{i=1}^k \bar{X}_i/k$, and zero proportions to the rest." Another type of impartial decision rule according to which the λ_j's are random variables will be described at the end of Example 1 in the next section. Now, it is (intuitively) clear that if the sample size n is indefinitely large, the rule $[\lambda_k \equiv 1]$, i.e., "Always assign the proportion 1 to the population

[5] It is unnecessary to specify here the exact relation between the statistics and the parameters. (a) the definition of the parameter which determines a distribution function $G(x, \theta)$ is more or less arbitrary, e.g., instead of writing $\theta = (b, c)$ we may write $\theta = (b^3/c, \cosh c)$, and (b) $D(\phi_1, \psi_1) = D(\phi_2; \psi_2)$, provided that $\phi_2 = f(\phi_1)$, $\psi_2 = g(\psi_1)$, where $f(x)$, $g(x)$ are strictly monotonic functions. It will be seen that Theorem 1 is invariant under such transformations of parameters and/or of statistics.

with the greatest $\bar{X}_i$", cannot be improved, no matter what the true parameter values may be. Our main result (Theorem 1) asserts that the statement is in fact valid for any n, provided that one restricts oneself to the class of impartial decision rules

In a similar way, if the X_{ij}'s are from normal populations having a common mean m and variances σ_i^2, D would be the class of all impartial decision rules which are based on the statistics

$$X_i = S_i^2 = \sum_{j=1}^{n} (X_{ij} - \bar{X}_i)^2/n - 1, \qquad i = 1, 2, \cdots, k,$$
$$Y = \sum_{i=1}^{k} \sum_{j=1}^{n} X_{ij}/kn,$$

and analogous remarks will apply to this case.

It should be observed that in a given case the appropriate statistics $\{X_i\}$; Y may not be as obvious as in the case of populations like the normal which admit sufficient statistics for their parameters This real difficulty is not to be confused with the ambiguities mentioned in footnote 4. Furthermore, given the X_i's there may not exist $Y = \psi(\{X_{ij}\})$ which is independent of the X_i's: we shall then assume, without invalidating our result, that the parameter which Y is supposed to estimate is known. Theorem 1 becomes operative only after such questions have been resolved.

5. The uniformly best decision rule. It is convenient to define here some terms which will be used subsequently without further explanation. All functions are assumed to be Borel measurable. Sets will be denoted by curly brackets: thus $\{f = c\}$ is the set on which $f = c$ holds, and $\{a_i\}$ is the set of all a_i in question. "Measure" will refer to ordinary Lebesgue measure in the xy plane.

DEFINITION 1. Given k independent random variables X_i, $i = 1, 2, \cdots, k$, such that each X has a frequency function, let $X_{(j)}$, $j = 1, 2, \cdots, k$, be the ordered set, $X_{(j)}$ being the jth X_i in ascending order of magnitude. Then $A_{ij} = \{X_i = X_{(j)}\}$, and a_{ij} is the characteristic function of the set A_{ij}, that is, $a_{ij} = 1$ for any point of A_{ij} and $= 0$ elsewhere.

Since the X_i's have a joint distribution which is absolutely continuous, the sets A_{ij} are well defined with probability one. Clearly, we have $\sum_{i=1}^{k} a_{ij} = 1$ for every j and $\sum_{j=1}^{k} a_{ij} = 1$ for every i, with probability one.

DEFINITION 2. Let $\beta = (b_1, b_2, \cdots, b_k)$ be a vector of real numbers b_i, and $\phi = (f_1, f_2, \cdots, f_k)$ a vector of real-valued functions $f_i(x)$ defined for every real x. We shall say that $\phi \in T(\beta)$ if for any $r, s = 1, 2, \cdots, k$ for which $b_r \leq b_s$, the set $\{f_r(x)f_s(y) < f_r(y)\, f_s(x), x < y\}$ is of measure zero.

We require the following

LEMMA. *Suppose that $X_1, X_2, \cdots, X_k$; Y are independent random variables, X_i having a frequency function $f_i(x)$ and that $\phi = (f_1, f_2, \cdots, f_k) \in T(\beta)$, where $\beta = (b_1, b_2, \cdots, b_k)$ with*

$$(1) \qquad b_1 \leq b_2 \leq \cdots \leq b_k.$$

Then, for any non-negative random variable $\lambda = \lambda(X_{(1)}, X_{(2)}, \cdots, X_{(k)}; Y)$ *and any* $p, q, m = 1, 2, \cdots, k$ *with* $p \leq q$, *we have*

$$\sum_{i=m}^{k} E(\lambda a_{ip}) \leq \sum_{i=m}^{k} E(\lambda a_{iq}). \tag{2}$$

PROOF. Since (2) holds trivially if $p = q$ or if $m = 1$ suppose that $p < q$ and $m \geq 2$. Writing $B(m, j) = \left\{\sum_{i=m}^{k} a_{ij} = 1\right\} = \sum_{i=m}^{k} A_{ij}$, (2) is equivalent to $\int_{B(m,q)} \lambda \, dP \geq \int_{B(m,p)} \lambda \, dP$, and hence to

$$\int_{B(m,q)B'(m,p)} \lambda \, dP \geq \int_{B(m,p)B'(m,q)} \lambda \, dP, \tag{3}$$

where B' denotes the complement of B, and P the probability measure in $(x_1, x_2, \cdots, x_k, y)$ space.

For any permutation $i_1 i_2 \cdots i_k$ of $123 \cdots k$, define $O(i_1 i_2 \cdots i_k) = A_{i_1 1} A_{i_2 2} \cdots A_{i_k k}$ Clearly, the O's corresponding to different permutations are disjoint and each of the sets $B(m, q)B'(m, p)$ and $B(m, p)B'(m, q)$ is the set-theoretic sum of certain O's. Now, it is easy to see that

$$\begin{aligned} &O \subset B(m, q)B'(m, p) \text{ if and only if } \begin{cases} i_p = 1, \text{ or } 2, \cdots, \text{ or } m - 1, \text{ and} \\ i_q = m, \text{ or } m + 1, \cdots, \text{ or } k. \end{cases} \\ &O^* \subset B(m, p)B'(m, q) \text{ if and only if } \begin{cases} i_p^* = m, \text{ or } m + 1, \cdots, \text{ or } k, \text{ and} \\ i_q^* = 1, \text{ or } 2, \cdots, \text{ or } m - 1. \end{cases} \end{aligned} \tag{4}$$

Hence a one-one correspondence between subsets $O(i_1 \cdots i_k)$ of $B(m, q)B'(m, p)$ and subsets $O^* = O(i_1^* \cdots i_k^*)$ of $B(m, p)B'(m, q)$ exists through interchange of the pth and qth elements of the defining permutations, the other elements remaining the same. It will be sufficient to prove that if O and O^* are any pair of corresponding subsets, the integral of λ over O is greater than or equal to its integral over O^*, for then (3) will follow by addition.

It is clear that for any O,

$$\begin{aligned} \int_{O(i_1 i_2 \cdots i_k)} \lambda \, dP &= \int_{\{x_{i_1} < x_{i_2} < \cdots < x_{i_k}\}} \lambda(x_{i_1}, x_{i_2}, \cdots, x_{i_k}; y) \\ &\qquad \cdot \left[\prod_{i=1}^{k} f_i(x_i) \, dx_i\right] dF(y) \\ &= \int_R \lambda(t_1, t_2, \cdots, t_k; y) \left[\prod_{r=1}^{k} f_{i_r}(t_r) \, dt_r\right] dF(y), \end{aligned} \tag{5}$$

where R is the domain $\{t_1 < t_2 < \cdots < t_k\}$ and $F(y)$ is the distribution function of Y. Let O and O^* be any pair of corresponding subsets. It follows from (5) that

$$\int_O \lambda \, dP - \int_{O^*} \lambda \, dP = \int_R Q \left[\prod_{r \neq p, q} f_{i_r}(t_r)\right] \prod_{r=1}^{k} dt_r \, dF(y),$$

where

(6) $$Q = \lambda(t_1, t_2, \cdots k_k; y)[f_{i_p}(t_p)f_{i_q}(t_q) - f_{i_q}(t_p)f_{i_p}(t_q)].$$

From (4) and (1) we have $b_{i_p} \leq b_{i_q}$. Since $p < q$ implies that $t_q < t_q$ over R, and $\phi \in T(\beta)$, it follows that the expression in square brackets in (6) is (except perhaps for a set of measure zero) non-negative over R Since λ is also non-negative, it follows that Q is non-negative over R, and the Lemma is proved.

We shall now state and prove the main result. Note that the statistic Y is not necessarily real-valued.

THEOREM 1. *Suppose that*

(A). *Ω is a given set of points $\omega = (\theta_1, \theta_2, \cdot \cdot, \theta_k)$. $\beta(\omega) = (b_1, b_2, \cdots, b_k)$ and $\gamma(\omega) = (g_1, g_2, \cdots, g_k)$ are defined for every ω such that $b_p \leq b_q$ implies $g_p \leq g_q$ for every $p, q = 1, 2, \cdots, k$.*

Given an $s = (p_1, p_2, \cdots, p_k)$ with $p_i \geq 0$ and $\sum_{i=1}^k p_i = 1$,

$$W(\omega, s) = \max_i \{g_i\} - \sum_{i=1}^k p_i g_i.$$

(B). *$X_1, X_2, \quad, X_k$; Y are independent random variables, each X_i having a frequency function $f(x, \theta_i) = f_i(x)$ say, and $\phi(\omega) = (f_1, f_2, \cdots, f_k)$.*

(C). *D is the class of all decision rules d such that*

$$d = d(X_{(1)}, X_{(2)}, \cdot \cdot, X_{(k)}; Y) = [\lambda_1, \lambda_2, \cdots, \lambda_k], \lambda_j \geq 0, \sum_{j=1}^k \lambda_j \equiv 1, \textit{and } s(d)$$
$$= (p_1(d), p_2(d), \cdot \cdot, p_k(d)) \textit{ where } p_i(d) = \sum_{j=1}^k \lambda_j a_{ij}, i = 1, 2, \cdots, k.$$

Given $d \in D$, $r(d \mid \omega) = E[W(\omega, s(d)) \mid \omega]$.

(D). *For every ω, $\phi \in T(\beta)$* [6]

Then, for every ω, $r(d_1 \mid \omega) = \sup_{d \in D} r(d \mid \omega)$ and $r(d_k \mid \omega) = \inf_{d \in D} r(d \mid \omega)$, where $d_1 \equiv [1, 0, 0, \cdots, 0]$ and $d_k \equiv [0, 0, \cdots, 0, 1]$.

COROLLARY. *Suppose that π_i, $i = 1, 2, \cdots, k$ are populations characterized by distribution functions $G(x, \theta_i) = h\left(\frac{x - b_i}{c_i}\right)$, $c_i > 0$. For any fixed x, let*

$$G(x \mid \omega, s) = \sum_{i=1}^k p_i G(x, \theta_i), \quad \text{and} \quad H(x \mid d, \omega) = E[G(x \mid \omega, s(d)) \mid \omega].$$

CASE 1. *If for every ω,* (i) $c_1 = c_2 = \cdots = c_k$,
(ii) $\phi \in T(\beta)$, *where* $\beta = (b_1, b_2, \cdots, b_k)$,

then, for every ω,

$$H(x \mid d_k, \omega) = \inf_{d \in D} H(x \mid d, \omega).$$

CASE 2. *If for every ω,* (i) $b_1 = b_2 = \cdots = b_k = b(\omega)$, *say,*
(ii) $\phi \in T(\beta)$, where $\beta = (c_1, c_2, \cdots, c_k)$,

[6] Note that $\phi \in T(-\beta)$ is equivalent to $\phi^* \epsilon T(\beta)$, where $\phi^* = (f_1^*, f_2^*, \cdot \cdot \cdot f_k^*)$, and f_i^* is the frequency function of $X_i^* = -X_i$.

then, for every ω,

$$H(x \mid d_1, \omega) = \begin{cases} \inf_{d \in D} H(x \mid d, \omega) & \text{if} \quad x < b(\omega), \\ \sup_{d \in D} H(x \mid d, \omega) & \text{if} \quad x > b(\omega). \end{cases}$$

PROOF. Choose and fix an arbitrary $\omega \in \Omega$. Without loss of generality we may assume the notation to be so chosen (by simultaneous interchanges of indices i in each of $\{\theta_i\}$, $\{b_i\}$, $\{g_i\}$, $\{p_i\}$, $\{X_i\}$, $\{f_i\}$, and $\{a_{ij}\}$, $j = 1, 2, \cdots, k$) that (1) holds. It then follows that $g_1 \leq g_2 \leq \cdots \leq g_k$ and we write

$$g_i = g_1 + h_1 + h_2 + \cdots + h_i, \qquad h_i \geq 0, \qquad i = 1, 2, \cdots, k. \tag{7}$$

Choose and fix an arbitrary member of the class of impartial decision rules, say $d = [\lambda_1, \lambda_2, \cdots, \lambda_k]$. We have

$$r(d \mid \omega) = \max_i \{g_i\} - \sum_{i,j=1}^{k} g_i E(\lambda_j a_{ij}). \tag{8}$$

Now

$$\begin{aligned} \sum_{i,j=1}^{k} g_i E(\lambda_j a_{ij}) &= \sum_{i,j=1}^{k} (g_1 + h_1 + \cdots + h_i) E(\lambda_j a_{ij}) \\ &= g_1 + \sum_{m,j=1}^{k} \left[\sum_{i=m}^{k} E(\lambda_j a_{ij}) \right] h_m . \end{aligned} \tag{9}$$

Since $\lambda_j = \lambda_j(X_{(1)}, X_{(2)}, \cdots, X_{(k)}; Y) \geq 0$, it follows from the Lemma that

$$\sum_{i=m}^{k} E(\lambda_j a_{ij}) \leq \sum_{i=m}^{k} E(\lambda_j a_{ik}) \quad \text{for every } m \text{ and every } j, \tag{10}$$

by writing $\lambda = \lambda_j$, $p = j$, and $q = k$ in (2) By using (7), (9) and (10) it follows that

$$\begin{aligned} \sum_{i,j=1}^{k} g_i E(\lambda_j a_{ij}) &\leq g_1 + \sum_{m,j=1}^{k} \left[\sum_{i=m}^{k} E(\lambda_j a_{ik}) \right] h_m \\ &= g_1 + \sum_{m=1}^{k} \sum_{i=m}^{k} h_m E(a_{ik}) \\ &= \sum_{i=1}^{k} g_i E(a_{ik}). \end{aligned} \tag{11}$$

Therefore, by (8) and (11),

$$r(d \mid \omega) \geq \max_i \{g_i\} - \sum_{i=1}^{k} g_i E(a_{ik}) = r(d_k \mid \omega), \tag{12}$$

by definition of d_k. The inequality $r(d \mid \omega) \leq r(d_1 \mid \omega)$ follows from (8) and (9) by a similar use of the Lemma. Since both $d \in D$ and $\omega \in \Omega$ are arbitrary, this completes the proof of Theorem 1.

The verification of the corollary is as follows Choose and fix an arbitrary x and write $h\left(\frac{x - b_i}{c_i}\right) = t_i(\omega)$.

CASE 1. Let $\gamma(\omega) = (1 - t_1, 1 - t_2, \cdots, 1 - t_k)$ Then $r(d \mid \omega) = H(x \mid d, \omega) - \min_i \{t_i\}$, and it follows from the Theorem that $H(x|d_1,\omega) = \sup_{d \in D} H(x \mid d, \omega)$ and $H(x \mid d_k, \omega) = \inf_{d \in D} H(x \mid d, \omega)$, for all ω.

CASE 2. Let $$\gamma(\omega) = \begin{cases} (t_1, t_2, \cdots, t_k) & \text{if } b(\omega) > x, \\ (1 - t_1, 1 - t_2, \cdots, 1 - t_k) & \text{otherwise.} \end{cases}$$

Then we have $$r(d \mid \omega) = \begin{cases} \max_i\{t_i\} - H(x \mid d, \omega) & \text{if } b(\omega) > x, \\ H(x \mid d, \omega) - \min_i \{t_i\} & \text{otherwise,} \end{cases}$$

so that $$H(x \mid d_1, \omega) = \begin{cases} \inf_{d \in D} H(x \mid d, \omega) & \text{if } b(\omega) > x, \\ \sup_{d \in D} H(x \mid d, \omega) & \text{otherwise,} \end{cases}$$

and conversely for $H(x \mid d_k, \omega)$, for all ω.

The preceding proofs suggest that perhaps (D) is not a necessary condition, but the following theorem for the case of two populations shows that it is indispensable if Theorem 1 is to hold in general.

THEOREM 2. *Suppose that (A), (B), and (C) hold with $k = 2$ and θ_1, θ_2 real-valued, that the set Ω of points $\omega = (\theta_1, \theta_2)$ is denumerable, that $\beta(\omega) = \omega$, that $g_1 \neq g_2$ for any ω, and that Y is a fixed constant. Let $\mu(\omega) = \min_i \{\theta_i\}$, $\nu(\omega) = \max_i \{\theta_i\}$, and defining the sets*

$$R(\omega) = \{f(t_1, \mu) f(t_2, \nu) < f(t_1, \nu) f(t_2, \mu), \quad t_1 < t_2\},$$
$$S(\omega) = \{f(t_1, \mu) f(t_2, \nu) > f(t_1, \nu) f(t_2, \mu), \quad t_1 < t_2\}$$

in the t_1, t_2-plane, put

$$R^*(t_1, t_2) = \sum_\omega R(\omega),$$
$$S^*(t_1, t_2) = \sum_\omega S(\omega).$$

Then a uniformly best decision rule in the class D exists if and only if the set R^*S^* is of measure zero. Subject to existence, the uniformly best rule, say d^*, may be defined as

$$d^* = \begin{cases} [1, 0] \text{ if } (X_{(1)}, X_{(2)}) \in R^*, \\ [0, 1] \text{ otherwise.} \end{cases}$$

The proof is quite simple, and will not be given. It is clear that under the hypotheses of this theorem, the conclusion of Theorem 1 is valid if and only if the set R^* is of measure zero, that is, if and only if condition (D) holds.

6. Examples and discussion. We begin with two applications of Theorem 1.

EXAMPLE 1 Suppose that grain is to be raised on a given area, say A, of land. k varieties, $\pi_1, \pi_2, \cdots, \pi_k$ say, are available, the yields per unit area being normally distributed with unknown means m_i and a common variance σ^2, also unknown. A preliminary field experiment (in which n plots of unit area were assigned to each variety) has been carried out, and $\{X_{ij}\}$, $j = 1, 2, \cdots, n$; $i = 1, 2, \cdots, k$ is the set of independent plot-yields obtained. The statistician is asked to suggest how the available land should be divided between the k varieties, the object being to make the total expected yield as large as possible.[7]

Suppose that an area Ap_i is assigned to π_i, $i = 1, 2, \cdots, k$, with $\sum_{i=1}^k p_i = 1$. Then the expected total yield is $\sum_{i=1}^k Ap_i m_i$. Our object is to choose the set $(p_1, p_2, \cdots, p_k) = s$ so as to minimize the "loss"

$$W(\omega, s) = \max_i \{Am_i\} - \sum_{i=1}^k Am_i p_i .$$

Since the m_i's are unknown, one must construct an appropriate s-valued function of the X_{ij}'s, say d, and set $s(d) = d(\{X_{ij}\})$. The expected "loss" in using this procedure is given by $E[(\omega, s(d)) \mid \omega] = r(d \mid \omega)$, and the problem is to construct a d which makes $r(d \mid \omega)$ as small as possible. (See (A) and (C). Here we have set $\theta_i = (m_i, \sigma)$, $\omega = (\theta_1, \theta_2, \cdots, \theta_k)$, $\beta(\omega) = (m_2, m_3, \cdots, m_k)$ and $\gamma(\omega) = (Am_1, Am_2, \cdots, Am_k)$).

Let $\bar{X}_i = \sum_{j=1}^n X_{ij}/n$, $i = 1, 2, \cdots, k$ and $S^2 = \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2/k(n-1)$.

Since $\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_k, S^2$ is a set of sufficient statistics, it is easy to see by taking conditional expectations that corresponding to any decision rule based on the X_{ij}'s, there exists one defined in terms of the $\bar{X}_i$'s and S^2 alone such that the risk functions r of the two are identically equal for all possible values of the unknown parameters. Clearly, one may confine oneself to decision rules of the type $d = s(\{\bar{X}_i\}; S^2)$. Now, the frequency function of $\bar{X}_i$ is $f_i(x) = (n/2\pi\sigma^2)^{1/2}. \exp[-n(x - m_i)^2/2\sigma^2]$, and it is readily seen that $m_r \leq m_s$ and $x < y$ imply $f_r(x)f_s(y) \geq f_s(x)f_r(y)$ It follows that in the class of all impartial procedures which are based on $\{\bar{X}_i\}$, S^2, the uniformly best procedure is to assign the whole area A to the variety with the greatest observed yield. (Note that by the corollary to Theorem 1, a much stronger result than the one required here holds. Cf. footnote 3)

Although Paulson did not set up a weight function in his discussion of the selection problem for the present case of samples of equal size from k normal populations having unknown means and a common variance, also unknown, he

[7] A double expectation is involved the expected consequence of a given decision, and the expected decision in using a particular decision rule. The argument given is justified since it is assumed that the random variables generated by the π's subsequent to decision are independent of the random variables on which decision is based. Cf. Section 3. This remark applies to Example 2 also

gave a class $\{d_c\}$ of decision rules and evaluated some probabilities ($P(G_1)$ and P_s^*. [1], pp. 96–97) which suggest that some of the applications he had in mind are similar to the one given here. In our notation, the rule d_c is defined as follows for any given $c \geq 0$.

$$d_c = [\lambda_1, \lambda_2, \cdots, \lambda_k], \qquad \text{where} \qquad \lambda_j = \left(Z_j \Big/ \sum_{j=1}^{k} Z_j\right), \qquad j = 1, 2, \cdots, k$$

$$\text{with} \qquad Z_j = \begin{cases} 1 & \text{if } X_{(k)} - c(S/\sqrt{n}) \leq X_{(j)} \leq X_{(k)}, \\ 0 & \text{otherwise.} \end{cases}$$

EXAMPLE 2. Suppose that a manufactured article has a numerical characteristic x, and a given article is "defective" if it has an $x < a$ and "acceptable" otherwise, where a is some constant. A consumer requires a large number (N) of articles, which can be supplied by each one of k manufacturers π_i, $i = 1, 2, \cdots, k$. The characteristic (say length) of articles produced by π_i is known to have a rectangular distribution with range from b to $b + c_i$, but the c_i's are not known. As a preliminary step, the consumer has obtained samples of ν articles from each manufacturer, and finds the corresponding lengths to be X_{ij}, $j = 1, 2, \cdots, \nu$, $i = 1, 2, \cdots, k$. The statistician is asked to suggest how the consumer should order a total of N articles from the k manufacturers.

If $a \leq b$, the number of defective articles received by the consumer will be zero no matter how the order is placed. Suppose therefore that $a > b$. Then, if n_i articles are ordered from π_i with $\sum_{i=1}^{k} n_i = N$, the expected number of defectives equals $N - \sum_{i=1}^{k} (n_i/N)\cdot g_i$, where $g_i = g(c_i)$ and $g(t)$ is given by

$$g(t) = \begin{cases} N\left(1 - \dfrac{a-b}{t}\right) & \text{if } t \geq a - b, \\ 0 & \text{otherwise.} \end{cases}$$

Writing $\beta(\omega) = (c_1, c_2, \cdots, c_k)$, $\gamma(\omega) = (g_1, g_2, \cdots, g_k)$, it is clear that the expected number of defectives is of the form $W(\omega, s) + h(\omega)$, where $h(\omega)$ is independent of $s = (n_1/N, n_2/N, \cdots, n_k/N)$, and W is defined as in (A)

We have now to consider what statistics X_i should be used to construct decision rules. Evidently, we are concerned with a "problem of the greatest c_i."

(a). Assuming $\nu > 1$, let $X_i = \max_j \{X_{ij}\} - \min_j \{X_{ij}\}$. Since the frequency function of X_i is $f_i(x) = \nu(\nu - 1)c_i^{-\nu}(c_i - x)x^{\nu-2}$ if $0 < x < c_i$ and zero elsewhere, it is a simple matter to show that $c_r \leq c_s$, $x < y$ imply $f_r(x)f_s(y) \geq f_s(x)f_r(y)$. It follows that in the class of all impartial rules which are based on the sample ranges, the uniformly best rule is to order all the N articles from the manufacturer with the greatest sample range.

(b). It may be objected that since the lower end points of all the distributions are the same, the use of sample ranges to construct decision rules is not particularly appropriate. Suppose therefore that one takes the statistics $X_i^* = \max_j \{X_{ij}\} - b$. The frequency function of X_i^* is $f_i^*(x) = \nu c_i^{-\nu}x^{\nu-1}$ for $0 < x < c_i$ and $= 0$ elsewhere, and as before, condition (D) holds. Hence the uniformly

best impartial procedure in this class is to order all the N articles from the manufacturer who supplied the article with the greatest length in the whole sample of $k\nu$ articles

It is important to observe that the uniformly best procedures according to (a) and (b) are not identical, and choosing between them is outside the scope of Theorem 1. Note also that the statistics X_i^* are sufficient for the c_i's. Therefore, corresponding to any decision rule there exists a decision rule which is defined in terms of the X_i^*'s and has the same risk function. In particular, there exists a decision rule in class (b) which is equivalent to the uniformly best impartial rule in class (a). It would be interesting to know whether this equivalent rule is also an impartial one

The two examples given above are purely illustrative, and the reader will readily construct others in which the statistician is faced with similar problems of decision. The second example does not, strictly speaking, belong to Case 2, and the reader is urged to consider some specific instances of this Case. There are various modifications of "the problem of the greatest one" which may be indicated here very briefly. These modifications are introduced by placing restrictions on the class of possible decisions. For example, in Example 1 the statistician may be required to select two or more varieties, and to assign proportions of the land to the varieties which he selects in such a way that no variety takes more than two-thirds of the available land. In that case, the uniformly best procedure (in the class of all impartial procedures which are based on the $\bar{X}_i$'s and S^2) would be to assign two-thirds of the land to the variety with the greatest observed mean yield, and the remainder to the variety with the next greatest. The proof is a slight elaboration of the proof of Theorem 1 and is left to the reader. Again, in Example 2 the consumer may wish to obtain all the articles which he requires from some one manufacturer. In that case, assuming that an impartial selection rule based on the X_i^*'s is to be used, it follows trivially from the case considered previously that the uniformly best procedure is to select the manufacturer with the greatest X_i^*. This is intuitively obvious, but the obvious requires proof (i.e. verification of (D)), as may be seen by turning to Example 3.

The intuitive notion referred to above is one which is employed quite frequently in practice. It may be described as follows. Let X_1 and X_2 be independent and similar estimates of unknown parameters m_1 and m_2, and suppose that in a given instance we have $X_1 > X_2$. "Then it is more reasonable to suppose that $m_1 > m_2$ than to suppose that $m_1 < m_2$." Theorem 2 shows that this notion is well-founded if and only if condition (D) is satisfied, with $\beta = (m_1, m_2)$. The condition states essentially that "the likelihood of the greater estimate corresponding to the greater parameter is always $\geq$ the likelihood of the contrary event," and it should be observed that X_1, X_2 being "good" estimates (e.g. maximum likelihood estimates) does not ensure that this will be the case. The following application of Theorem 2 is an illustration of these remarks.

EXAMPLE 3. Suppose that π_i, $i = 1, 2$ are Cauchy-type populations having medians m_i, and that the set Ω of possible points $\omega = (m_1, m_2)$ consists of just

the two points $\omega_1 = (1, -1)$ and $\omega_2 = (-1, 1)$. X_1 and X_2 are single observations from the two populations, and the statistician is required to decide which population has the greater median.

Here it would be reasonable for the statistician to use a decision rule, say d^*, which minimizes $r(d \mid \omega) = P(\text{incorrect decision} \mid \omega, d)$, where "$\pi_1$ has the greater median" and "π_2 has the greater median" are the two possible decisions. That this risk function is included in the scheme described by (A) and (C) may be seen as follows. Let the only admissible values of s be (1, 0) and (0, 1), corresponding to the decisions "$m_1 > m_2$" and "$m_1 < m_2$" respectively, and setting $\beta(\omega) = (m_1, m_2)$, define $\gamma(\omega_1) = (1, 0)$, $\gamma(\omega_2) = (0, 1)$. Then for any d such that $s(d)$ equals (1, 0) or (0, 1) only, the expected value of W is for either ω the probability of error in using the rule d

Now, if $d = d(X_{(1)}, X_{(2)}) = [\lambda_1, \lambda_2]$ is any impartial decision rule, it will equal either [1, 0] or [0, 1], corresponding to the decisions "the population with the greater X has the smaller median" and "the population with the greater X has the greater median" respectively. Since the frequency function of X_i is $f_i(x) = 1/\pi[1 + (x - m_i)]^2$, a little calculation shows that in the class of impartial decision rules a uniformly best one exists, and is given by

$$d^* = \begin{cases} [1, 0] & \text{if } X_{(1)}X_{(2)} > 2, \\ [0, 1] & \text{otherwise.} \end{cases}$$

In conclusion, we remind the reader that although the weight function W defined according to (A) is general enough to include all problems of the type considered in this paper, the sampling scheme as also the class of decision rules to which our results apply is very limited. We have (i) assumed that the samples from the k populations are all of the same size, and (ii) given no objective criterion for choosing appropriate statistics, and no justification for the use of impartial decision rules based on these "appropriate statistics." In view of the applications, it would be of interest to extend the general argument of this paper to the numerous situations where Theorem 1 does not apply or is otherwise unsuitable.

The problem of selection was suggested to the author by Professor Hotelling. The author would like to acknowledge his indebtedness also to Professor Robbins. This paper could not have been written without his constant encouragement and helpful suggestions.

REFERENCES

[1] Edward Paulson, "A multiple decision procedure for certain problems in analysis of variance," *Annals Math. Stat.*, Vol. 20 (1949), pp 95–98.

[2] Frederick Mosteller, "A k-sample slippage test for an extreme population," *Annals Math Stat*, Vol 19 (1948), pp 58–65

[3] Abraham Wald, *On the Principles of Statistical Inference*, Notre Dame Mathematical Lectures, No. 1, (1942), Notre Dame, Indiana.

[4] Herbert Robbins, "Mixture of distributions," *Annals Math Stat.*, Vol. 19 (1948), pp. 360–369

COMPLETENESS IN THE SEQUENTIAL CASE

By E. L. Lehmann and Charles Stein

University of California, Berkeley

1. Summary. Recently, in a series of papers, Girshick, Mosteller, Savage and Wolfowitz have considered the uniqueness of unbiased estimates depending only on an appropriate sufficient statistic for sequential sampling schemes of binomial variables. A complete solution was obtained under the restriction to bounded estimates. This work, which has immediate consequences with respect to the existence of unbiased estimates with uniformly minimum variance, is extended here in two directions. A general necessary condition for uniqueness is found, and this is applied to obtain a complete solution of the uniqueness problem when the random variables have a Poisson or rectangular distribution. Necessary and sufficient conditions are also found in the binomial case without the restriction to bounded estimates. This permits the statement of a somewhat stronger optimum property for the estimates, and is applicable to the estimation of unbounded functions of the unknown probability

2. Introduction. The notions of completeness and bounded completeness of a family of distributions were introduced in [1, 2] in connection with the problems of similar regions and unbiased estimation. The question of whether either of these two properties pertains to various families of distributions that are of interest in statistics was discussed in [2] under the assumption of fixed sample size. The only sequential problems of this kind that have been treated in the literature (with quite different terminology) refer to the binomial case. For this case Girshick, Mosteller and Savage [3] found necessary (and also certain sufficient) conditions on the sequential sampling scheme for completeness, while Wolfowitz [4] and Savage [5] gave necessary and sufficient conditions for bounded completeness.

If T is a random variable distributed over an additive class of sets in some space according to a distribution P_θ^T with θ in some set ω, then the family $\mathcal{P}^T = \{P_\theta^T \mid \theta \in \omega\}$ of possible distributions of T is said to be complete if

$$\int f(t)\, dP_\theta^T(t) = 0, \qquad \text{for all} \quad \theta \in \omega, \tag{1}$$

implies

$$f(t) = 0, \qquad \text{a.e. } \mathcal{P}^T, \tag{2}$$

that is, for all t except possibly in a set N for which $P_\theta^T(N) = 0$ for all $\theta \in \omega$. The family $\mathcal{P}^T$ is said to be boundedly complete if this implication holds under the assumption that f is bounded.

The relation of these concepts to the problem of unbiased estimation is an

immediate consequence of a theorem of Blackwell [6]. Let X be a random variable with distribution P_θ^X, $\theta \epsilon \omega$, and let T be a sufficient statistic for θ Denote by P_θ^T the distribution of T, and suppose that $\mathcal{P}^T$ is complete. Then every function $g(\theta)$ for which there exists an unbiased estimate, that is, a function ϕ such that

$$E_\theta \phi(X) = g(\theta), \qquad \text{for all} \quad \theta \epsilon \omega,$$

possesses an unbiased estimate with uniformly minimum variance. One can say furthermore that if $\phi(X)$ is any unbiased or bounded unbiased estimate of $g(\theta)$, then the optimum estimate guaranteed by the above statements is the conditional espectation of $\phi(X)$ given T.

The aim of the present paper is to obtain certain results concerning completeness in sequential sampling schemes. Some necessary conditions for completeness are given in section 3, and these are used to obtain necessary and sufficient conditions for completeness when the random variable being sampled has a Poisson or rectangular distribution In section 4 it is shown that certain necessary conditions given in [3] for the binomial case are also sufficient

3. A necessary condition for completeness. The sequential sampling schemes with which we are concerned are of the following nature. There is given a sequence of real valued random variables $X_1, X_2, \cdots$ with a joint distribution depending on a real parameter θ, which ranges over a set ω We shall assume that for each m the set of variables $X_1, \cdots, X_m$ admits a real valued sufficient statistic $T_m = t_m(X_1, \cdot \cdot, X_m)$ for θ, and that for each m the family $\mathcal{P}^{T_m}$ of distributions of T_m is complete. We next suppose that there is given a stopping rule, which is such that after m observations have been taken, the decision of whether or not to take an $m+1st$ observation depends only on the value of $t_m(X_1, \cdot \quad, X_m)$ It follows (see [6]) that if the total number of observations is n (a random variable which may be infinite), then (T_n, n) is a sufficient statistic for θ. We shall say that the sequential procedure is complete if the family of distributions of (T_n, n) is complete. Throughout, we shall assume that all sequential procedures in question are closed, i.e. that for each $\theta \epsilon \omega$, n is finite with probability 1.

Let Y be a random variable distributed over a Euclidean space according to a distribution P_θ^Y with θ in ω. We shall say that a point y lies in the positive sample space of Y if there exists $\theta \epsilon \omega$ such that every open set containing y has positive probability for this θ, and that y is an impossible point if it lies in the complement of the positive sample space. Consider now a sequential sampling scheme as described above. For any integers $m < p$ we shall denote by W_p^m the positive sample space of T_p given the first m steps of the stopping rule, that is, given for $i = 1, \cdots, m$ the set S_i of values of T_i for which sampling is discontinued after the ith observation Since all the T's are real valued, the sets W_p^m are sets of real numbers satisfying the obvious condition $W_p^{m-1} \geqq W_p^m$. The union $\bigcup S_m$ (S_m is the set of points of W_m^{m-1} for which no $m+1st$ observation is

taken) will be called the set of stopping or boundary points, the points belonging to some $W_m^{m-1} - S_m$ are the continuation points.

We need the following

LEMMA 1. *A necessary condition for a sequential procedure of the type described above to be complete is that every procedure obtained from the given one by truncation be complete.*[1]

This is an immediate consequence of the following more general

LEMMA 2. *Let $X_1, X_2, \cdots$ be as before a sequence of random variables such that for each m the set $X_1, \cdots, X_m$ admits a real valued sufficient statistic $T_m = t_m(X_1, \cdots, X_m)$. Let $\Sigma_1, \Sigma_2, \cdots, \Sigma_r$ each be a complete, closed, sequential procedure based on these sufficient statistics. Let $\Sigma_1 \cup \Sigma_2 \cup \cdots \cup \Sigma_r$ denote the sequential procedure according to which we continue taking observations until at least one of the stopping rules $\Sigma_1, \cdots, \Sigma_r$ tells us to stop. Then the procedure $\Sigma_1 \cup \cdots \cup \Sigma_r$ is complete*

This clearly implies Lemma 1. For if one takes for Σ_1 any closed, complete sequential procedure and for Σ_2 a procedure of fixed sample size, then $\Sigma_1 \cup \Sigma_2$ is the associated truncated procedure.

PROOF OF LEMMA 2 It is sufficient to prove the result for the case $r = 2$. Let n_1, n_2, n denote the number of observations taken under $\Sigma_1, \Sigma_2, \Sigma_1 \cup \Sigma_2$ respectively. Then $n = n_1$ if $n_1 \leqq n_2$, $n = n_2$ if $n_1 \geqq n_2$. Let f be any function on $\Sigma_1 \cup \Sigma_2$ such that

$$E_\theta f(T_n, n) = 0 \quad \text{for all} \quad \theta \in \omega.$$

Then

$$\left.\begin{aligned} E_\theta E[f(T_n, n) \mid T_{n_1}, n_1] &= 0 \\ E_\theta E[f(T_n, n) \mid T_{n_2}, n_2] &= 0 \end{aligned}\right\} \quad \text{for all} \quad \theta \in \omega.$$

Since Σ_1 and Σ_2 are complete it follows that

$$E[f(T_n, n) \mid T_{n_1} = t_1, n_1 = \gamma_1] = E[f(T_n, n) \mid T_{n_2} = t_2, n_2 = \gamma_2] = 0, \text{ a.e.}$$

Hence

$$\begin{aligned} 0 = {} & P(n_1 \leqq n_2 \mid T_{n_1} = t_1, n_1 = \gamma_1) f(t_1, \gamma_1) \\ & + P(n_1 > n_2 \mid T_{n_1} = t_1, n_1 = \gamma_1) E[f(T_{n_2}, n_2) \mid T_{n_1} = t_1, n_1 = \gamma_1, n_1 > n_2], \end{aligned} \tag{3}$$

and the analogous condition holds with the subscripts 1 and 2 interchanged.

We shall prove that $f(T_n, n) = 0$, a.e., by induction over the possible values of n. Suppose, therefore, that for some integer m

$$P_\theta(n \leqq m, f(T_n, n) \neq 0) = 0.$$

(This is certainly true for $m = 0$.) It then follows that if we take $\gamma_1 = m + 1$ in (3) the second term of the right hand side vanishes, so that

$$0 = P(n = n_1 \mid T_{n_1} = t_1, n_1 = m + 1) f(t_1, m + 1).$$

[1] The authors would like to thank Mr E. Fay for pointing out an error in the original proof of this Lemma.

Hence,

$$P_\theta(n = n_1 = m + 1, f(T_{n_1}, n_1) \neq 0)$$
$$\leqq P_\theta(n = n_1 = m + 1, P(n = n_1 \mid T_{n_1}, n_1) = 0) = 0$$

Analogously we see that

$$P_\theta(n = n_2 = m + 1, f(T_{n_2}, n_2) \neq 0) = 0$$

and, adding, that

$$P_\theta(n = m + 1, f(T_n, n) \neq 0) = 0.$$

This completes the induction.

We need further the notion of strong completeness. Consider a random variable $W = (U, V)$, suppose that the distribution of W depends on θ, and that U is a sufficient statistic for θ Let P_u^V be the conditional distribution of V given $U = u$—this is independent of θ since U is a sufficient statistic for θ—and let $\mathcal{P}^{V*} = \{\mathcal{P}_u^V\}$. We say that the pair $\mathcal{P}^W$, $\mathcal{P}^{V*}$ is strongly complete if the conditions
(i) $E_\theta f(V)$ exists for all θ,
(ii) $E(f(V) \mid U = u) = 0$ for almost all u,
imply

$$f(v) = 0, \qquad \text{a.e. } \mathcal{P}^V.$$

For brevity, we shall then usually say that $\{\mathcal{P}_u^V\}$ is strongly complete.

We can now state the following necessary condition for completeness.

THEOREM. *If a closed sequential procedure of the type considered above is complete, then*

(i) *S_m is almost empty for every m for which $W_{m+1}^{m-1} - W_{m+1}^m$ is almost empty,*

(ii) *for each m for which S_m is not almost empty, the family of conditional distributions of T_m given $T_{m+1} = t$ (as t ranges over $W_{m+1}^{m-1} - W_{m+1}^m$) is strongly complete.*

PROOF. For any $t \in W_{m+1}^{m-1} - W_{m+1}^m$ the positive sample space of T_m given $T_{m+1} = t$ is clearly contained in S_m. Suppose first that (ii) is violated and consider the sequential procedure obtained from the given one by truncation after $m + 1$ observations By the lemma it will be enough to show that the truncated procedure is not complete. For this purpose let us assume that regardless of the stopping rule all $m + 1$ variables $X_1, \cdots, X_{m+1}$ are observed. We want to construct an estimate of zero based on the sufficient statistic for the truncated procedure. This estimate must be a function of T_1 for $T_1 \in S_1$, of T_2 for $T_2 \in S_2$, etc. That is, although we may imagine that the full sample of size $m + 1$ is taken, we must be careful not to use observations that are impossible when the stopping rule is followed.

We shall now show that there exists an unbiased estimate of zero which is zero over $S_1, \cdots, S_{m-1}$, equal to $f(T_m)$ on S_m and $g(T_{m+1})$ on W_{m+1}^m where f and g will be defined below. Since expectation equals expectation of conditional expectation, a statistic is an unbiased estimate of zero if its expectation exists

and its conditional expectation given $T_{m+1} = t$ is zero for almost all t. In our case this condition is equivalent to

$$(4) \qquad \int_{S_m} f(u)\, dP_m(u \mid T_{m+1} = t) + g(t) \int_{W_m^{m-1} - S_m} dP_m(u \mid T_{m+1} = t) = 0$$

for almost all $t \in W_{m+1}^m$,

$$(5) \qquad \int_{S_m} f(u)\, dP_m(u \mid T_{m+1} = t) = 0$$

for almost all $t \notin W_{m+1}^m$, i.e. for almost all $t \in W_{m+1}^{m-1} - W_{m+1}^m$, since $t \notin W_{m+1}^{m-1}$ implies $P(S_m | T_{m+1} = t) = 0$,

together with the existence of $E_\theta(f(T_m) \mid n = m)$ and $E_\theta(g(T_{m+1}) \mid n = m + 1)$ Since (ii) does not hold there exists f not vanishing a.e such that $E_\theta(f(T_m) \mid n = m)$ exists and (5) is satisfied. If g is defined by (4), $E_\theta(g(T_{m+1}) \mid n = m + 1)$ exists, and this completes the proof of the necessity of (ii).

The necessity of (i) is now obvious For if (i) is violated, then (5) is satisfied vacuously, and we can take f to be an arbitrary positive valued function (for example) and (4) will then be satisfied.

As immediate consequences of this theorem we shall obtain two conditions, which are easier to apply than condition (ii).

COROLLARY 1 *A necessary condition for completeness is that for no m there exists a subset A of S_m such that*

$$P_\theta(A) > 0 \quad \textit{for some} \quad \theta$$

and

$$P(A \mid T_{m+1} = t) = 0 \quad \textit{for almost all} \quad t \in W_{m+1}^{m-1} - W_{m+1}^m .$$

COROLLARY 2. *Suppose that the sequence of X's is such that in the non-sequential case for all m, p with $m < p$ the positive sample space of T_m given $T_p = t$ is the intersection of the unconditional positive sample space of T_m with the interval $[0, t]$. Then a necessary condition for a sequential procedure to be complete is that each S_m differ from a half-open interval (possibly empty) $[a_m, b_m)$ with $a_m \leqq b_m$, $a_1 = 0$, $a_{m+1} = b_m$, by a set of probability 0.*

PROOF. Let r be the first value of m for which this condition is not satisfied. Then there exists $c > b_{r-1}$ such that the sets $S_r \cap [c, \infty)$ and $\bar{S}_r \cap [b_{r-1}, c)$ both have positive probability The result now follows from Corollary 1 if one puts $A = S_r \cap [c, \infty)$.

Next we consider some examples.

EXAMPLE 1. Let $X_1, X_2, \cdots$ be independently normally distributed with known variance and unknown mean θ. In this case $T_m = \sum_{i=1}^m X_i$, and since the positive sample space of T_{m+1} is the infinite interval regardless of the values of $T_1, \cdots, T_m$ it follows from condition (i) of the theorem that no sequential procedure is complete, with the trivial exception of the procedures with fixed sample size

EXAMPLE 2. Let $X_1, X_2, \cdots$ be independently uniformly distributed over the interval $(0, \theta)$, $0 < \theta < \infty$. Then $T_m = \max(X_1, \cdots, X_m)$ and Corollary 2 gives a necessary condition for completeness If the procedure is truncated we can deduce sufficiency of this condition from (5) However, this proof does not apply to the general case The following proof of sufficiency is similar to some of the proofs in [3, 4, 5].

Suppose $S_1, S_2, \cdots$ form a set of adjoining intervals (some of them possibly empty), $S_m = [a_m, b_m)$, and suppose there is a non-zero unbiased estimate of zero, $\Phi = \phi(T_n, n)$. Let m be the smallest integer for which ϕ is not zero almost everywhere on S_m Then

$$E_\theta(\Phi) = P_\theta(n = m)E_\theta(\Phi \mid n = m) + \sum_{j=m+1}^{\infty} P_\theta(n = j)E_\theta(\Phi \mid n = j) \overset{(\theta)}{\equiv} 0,$$

and hence

$$P_\theta(n = m)E_\theta(\Phi \mid n = m) \overset{(\theta)}{\equiv} -\sum_{j=m+1}^{\infty} P_\theta(n = j)E_\theta(\Phi \mid n = j). \tag{6}$$

Now the right hand side of (6) is zero when $\theta \leqq b_m$, since it is then impossible that $T_j \in S_j$ for any $j > m$ Hence

$$E_\theta[\phi(T_m, m) \mid a_m \leqq T_m < b_m] = 0 \quad \text{for all} \quad \theta \leqq b_m,$$

and therefore

$$\int_{a_m}^{\theta} \phi(x, m)x^{m-1}\, dx = 0 \quad \text{for all} \quad \theta \text{ in } [a_m, b_m].$$

But this implies $\phi(x, m) = 0$ almost everywhere in S_m, which is a contradiction.

EXAMPLE 3 Let $X_1, X_2, \cdots$ be independently distributed according to a Poisson distribution with mean θ. Then $T_m = \sum_{i=1}^m X_i$ and again we can apply Corollary 2. To prove sufficiency we proceed as in example 2. If the condition of Corollary 2 is satisfied we may write without ambiguity $\psi(T_n)$ for $\phi(T_n, n)$. Let c be the smallest value of T_n for which $\psi(T_n) \neq 0$. Then if the probability of $T_n = j$ is $k(j)\theta^j e^{-\theta m_j}$, the identity $E_\theta(\Phi) \overset{(\theta)}{\equiv} 0$ implies

$$\phi(c)k(c)\theta^c\rho \overset{(\theta)}{\equiv} \sum_{j=c+1}^{\infty} \phi(j)k(j)\phi^j \cdot \varphi(c)k(c)\theta^c e^{-\theta m_c} \equiv \sum_{j=c+1}^{\infty} \varphi(j)k(j)\theta^j e^{-\theta m_j}.$$

Dividing this equation by θ^c and letting θ tend to zero we see that the right hand side tends to zero, which implies $\phi(c) = 0$ and hence a contradiction.

4. The binomial case. As was mentioned in section 1, the problem of bounded completeness was solved for the binomial case in [3, 4, 5]. Since presumably one is unwilling to estimate the bounded parameter p by means of an unbounded estimate, further work here may seem unnecessary. However, the problem of completeness seems to be of interest for two reasons. If the procedure is bound-

edly complete without being complete then, even though one may be reluctant to use such an estimate, there may exist an unbounded unbiased estimate of p, which for some values of p has smaller variance than the minimum variance bounded estimate. (An example of this is given in [2]). Since this possibility is ruled out when the procedure is complete it is seen that completeness permits statement of a stronger optimum property. Apart from this one may be interested in estimating some unbounded function of p such as $1/p$. In this case bounded completeness does not permit any statements concerning existence of optimum estimates

In the present section we shall change our notation somewhat. We are concerned with a sequence of independent trials with constant probability p of success. On the basis of m trials the total number y of successes is a sufficient statistic for p. Instead of representing the sufficient statistic for the sequential procedure by (y, n), we shall use the representation (x, y) where x is the total number of failures, so that $x + y = n$. The couples (x, y) may be thought of as making up the points with integral-valued coordinates of the first quadrant of an xy-plane, and as before may be classified as boundary points, continuation points, and impossible points. Adopting the terminology of [3], we shall call the value of $x + y$ the index of the point (x, y), so that the points of index m lie on the line $x + y = \mathrm{m}$.

Girshick, Mosteller and Savage defined a sequential procedure to be simple if for each m the continuation points of index m form an interval They proved that a necessary and sufficient condition for a bounded procedure to be complete is that it be simple. (A procedure is said to be bounded if there exists N so that the number of observations is $\leqq N$.) They also showed that in general simplicity is not sufficient for completeness. However, it was shown later [4, 5] that simplicity is sufficient for bounded completeness

A sequential procedure is said to be closed if the probability of termination is unity for every p with $0 < p < 1$. It was proved by Girshick, Mosteller and Savage that a necessary condition for completeness of a closed sequential procedure is that no procedure obtained from the given one by removing a boundary point be closed. (Removing a boundary point here means converting it into a continuation point.) We shall prove below that this condition together with simplicity is also sufficient for completeness. An interesting question is whether these two conditions are sufficient for completeness for the general sequential schemes considered in section 2, when simplicity is replaced by the condition that every procedure obtained from the given one by truncation is complete, and when the second condition is modified by the appropriate null set qualifications. It is easily seen that both of these conditions are necessary.

The following definitions will be needed below. A boundary point (a, b) is a lower (upper) boundary point if for some $x < 0$ (>0) the point $(a + x, b - x)$ is a continuation point. An impossible point (a, b) is a lower (upper) impossible point if for some $x < 0$ (>0) the point $(\mathrm{a} + x, b - x)$ is either a continuation point or a boundary point.

If the procedure is unbounded every boundary point is either a lower or an upper boundary point. If it is simple, no point can be both an upper and a lower boundary point. The same remarks apply to impossible points.

THEOREM *A necessary and sufficient condition for completeness of a closed procedure in the binomial case is that*
(i) *the procedure is simple,*
and
(ii) *the removal of any boundary point destroys closure*

PROOF Necessity was proved in [3] as was sufficiency for bounded procedures. Sufficiency for unbounded procedures will follow from the following two facts, which we shall prove below.

I. Suppose (i) holds and there exist numbers a, $M > 0$ such that for all boundary points (x, y) of index $m \geqq M$ the ratio $y/x \geqq a$. Let $f(x, y)$ be a non-zero unbiased estimate of zero defined over the set B of boundary points, and let m_0 be the smallest index for which there are points with $f(x, y) \neq 0$ Then $f(x, y) = 0$ for all lower boundary points of index m_0.

II. If (i) holds and if for every positive number a there exist infinitely many boundary points (x, y) with $y/x \leqq a$, then one may remove any lower boundary point without destroying closure

Suppose now that a sequential procedure satisfies (i) and (ii). Then, since no lower boundary point can be removed without destroying closure, it follows from II. that there exist a and M such that $y/x \geqq a$ for all boundary points of index $\geqq M$. Hence if $f(x, y)$ is an unbiased estimate of zero, and if m_0 is defined as in I., $f(x, y) = 0$ for all lower boundary points of index m_0. Because of symmetry the statements concerning upper boundary points analogous to I. and II also hold. It then follows analogously that $f(x, y) = 0$ for all upper boundary points of index m_0. But for a simple unbounded procedure every boundary point is either an upper or a lower boundary point, and hence we obtain a contradiction with the definition of m_0.

Before proving I. and II. we state the following corollary, which generalises an example given in [3].

COROLLARY. *A sequential procedure that is not bounded and that has a finite non-zero number of lower boundary points is not complete. The analogous result holds for upper boundary points.*

PROOF OF COROLLARY. This follows easily from II., since if a procedure of this type is to be closed there must exist for each $a > 0$ infinitely many upper boundary points (x, y) with $y/x \leqq a$.

In the remainder of the paper we are concerned with the proofs of I. and II.

PROOF OF I Assume I to be false, and let (x_0, y_0) be the lowest boundary point of index m_0 for which $f(x_0, y_0) \neq 0$ Then $y > y_0$ for all other boundary points (x, y) for which $f(x, y) \neq 0$. Hence if the probability of a point (x, y) is $c(x, y)p^y q^x$ and if $k(x, y) = c(x, y)f(x, y)$,

$$k(x_0, y_0)p^{y_0}q^{x_0} \equiv -\Sigma k(x, y)p^y q^x,$$

where the summation extends over all boundary points of index $\geqq m_0$ for which $y > y_0$. Dividing both sides by p^{y_0} we see that

$$k(x_0, y_0)q^{x_0} \equiv -p\Sigma k(x, y)p^{y-y_0-1}q^x.$$

If we can show that the expression multiplying $-p$ on the right hand side remains bounded as p tends to zero, we have a contradiction. For letting p tend to zero, we would then see that the right hand side tends to zero and the left hand side to $k(x_0, y_0)$, and hence that $f(x_0, y_0) = 0$.

To prove this, note that

$$| \Sigma k(x, y)p^{y-y_0-1}q^x | \leqq \Sigma | k(x, y) | p^{y-y_0-1}.$$

The right hand side is a power series in p. We shall show that this series converges for some $p_0 > 0$. This implies uniform convergence for $| p | < p_0$, and therefore the series remains bounded at $p = 0$ By assumption there exist numbers a and M' such that $y/x \geqq a$ for all boundary points with $y > M'$. From now on we shall consider all series as being summed over the set of boundary points for which $y > M'$ and hence $q^x \geqq q^{y/a}$. Since only a finite number of terms are omitted this does not affect any convergence properties.

Let $0 < p_1 < 1$. Then, since f is an unbiased estimate of zero, the series

$$\Sigma k(x, y)p_1^y q_1^x$$

converges absolutely. Hence, so does

$$\Sigma | k(x, y) | p_1^{y-y_0-1}q_1^{x-\frac{1}{a}(y_0+1)} \geqq \Sigma | k(x, y) | (q_1p_1^{\frac{1}{a}})^{y-y_0-1} = \Sigma | k(x, y) | p_0^{y-y_0-1},$$

and consequently the last series is convergent.

Proof of II. Let R be any closed simple procedure satisfying the conditions of II., and let (x_0, y_0) be any lower boundary point of R We denote by R^* the procedure obtained from R by taking (x_0, y_0) to be a continuation point and by n^* the number of observations for R^*.

We first prove that any upper impossible point of R is also an impossible point of R^*. The negation of this would imply that one can get from a lower boundary point to an upper impossible point going only through impossible points. This would require at least one step of either of the following kinds:

Lower impossible point $\rightarrow$ upper impossible point,

Lower boundary point $\rightarrow$ upper impossible point.

One can easily convince oneself with the aid of a diagram that any procedure under which such steps are permitted cannot be simple.

Let $0 < p, \pi < 1$, and let a be such that $0 < a < p/q$. If p is the true probability of success, y/x tends in probability to p/q, and hence there exists N such that

$$P(y/x \geqq a \mid p) > \pi$$

whenever the index of (x, y) exceeds N. By assumption there exists $N_1 > N$ and a boundary point (x_1, y_1) of R^* of index N_1 such that $y_1/x_1 \leqq a$. Then the

probability exceeds π that the random point (x, y) of index N_1 will lie above (x_1, y_1). Since (x_1, y_1) is a boundary point, the probability is therefore greater than π that the point (x, y) of index N is either an upper impossible point for R and hence impossible for R^*, or a stopping or continuation point for R. We have therefore proved that the probability is $>\pi$ that either $n^* \leqq N_1$ or the point (x, y) of index N_1 is a continuation point of R.

But given that one has reached a continuation point (a, b) of R, there exists N_2 such that

$$P(n^* \leqq N_2 \mid p, (a, b)) \geqq \pi.$$

For

$$P(n^* > N_2 \mid (a, b)) = P(n > N_2 \mid (a, b)) \to 0 \quad \text{as} \quad N_2 \to \infty.$$

Since there are only a finite number of continuation points of index N_1, it is now clear that there exists N_0 such that

$$P(n^* \leqq N_0 \mid p) \geqq \pi + \pi^2 - 1,$$

which can be made arbitrary close to 1 by proper choice of π. Therefore R^* is closed.

REFERENCES

[1] E. L. Lehmann and H. Scheffé, "On the problem of similar regions," *Proc Nat Acad. Sci.*, Vol. 33 (1947), pp. 382–386.

[2] E. L. Lehmann and H. Scheffé, "Completeness, similar regions and unbiased estimation," unpublished

[3] M. A Girshick, Frederick Mosteller and L. J. Savage, "Unbiased estimates for certain binomial sampling problems, with applications," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13–23

[4] J. Wolfowitz, "On sequential binomial estimation," *Annals of Math. Stat*, Vol. 17 (1946), pp. 489–493.

[5] L. J Savage, "A uniqueness theorem for unbiased sequential binomial estimation," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 295–297.

[6] D. Blackwell, "Conditional expectation and unbiased sequential estimation," *Annals of Math. Stat*, Vol. 18 (1947), pp 105–110.

SOME ESTIMATES AND TESTS BASED ON THE r SMALLEST VALUES IN A SAMPLE

By John E. Walsh[1]

The Rand Corporation

1. Summary. Let us consider a situation where only the r smallest values of a sample of size n are available. This paper investigates the case where n is large and r is of the form $pn + O(\sqrt{n})$.

Properties of some well known non-parametric point estimates, confidence intervals and significance tests for the $100p\%$ point of the population are investigated. If the sample is from a normal population, these non-parametric estimates and tests have high efficiencies for small values of p (at least 95% if $p \leqq 1/10$).

The other results of the paper are restricted to the special case of a normal population. Asymptotically "best" estimates and tests for the population percentage points are derived for the case in which the population standard deviation is known. For the case in which the population standard deviation is unknown, asymptotically most efficient estimates and tests can be obtained for the smaller population percentage points by suitable choice of p and $O(\sqrt{n})$.

The results derived have application in the field of life testing. There the variable associated with an item is the time to failure and the r smallest sample values can be obtained without the necessity of obtaining the remaining values of the sample. By starting with a larger number of units but stopping the experiment when only a small percentage of the units have "died", it is often possible (using the results of this paper) to obtain the same amount of "information" with a substantial saving in cost and time over that which would be required if a smaller number of units were used and the experiment conducted until all the units have "died". Jacobson called attention to applications of this type in [1].

2. Introduction and statement of results. In life testing, information concerning the smaller population percentage points may be of primary interest. The principal aim of this paper is to investigate the properties of some well known non-parametric estimates and tests of the smaller population percentage points which are based on statistics of the type used for the sign test. These non-parametric results are easy to apply and have several other desirable properties (see Theorem 1 and its discussion). In particular, if the $100p\%$ point is to be investigated, it is only necessary to fail approximately $100p\%$ of the number of starting items to obtain the required statistics (n large). Thus, if the non-parametric results should also happen to be reasonably efficient, they

[1] The author would like to express his appreciation to Max Halperin for calling attention to this problem and for valuable advice and assistance in the preparation of the paper.

would appear to be ideal for a life testing situation where a smaller population percentage point is to be investigated.

Examination shows that life tests of the "wear out" type sometimes yield empirical distributions which are approximately normal. Also in many cases an approximately normal distribution can be obtained by an appropriate monotonic change of variable. Thus the case in which the n observations are a sample from a normal population will receive special consideration in this paper

Investigation of the efficiency of the non-parametric estimates and tests will be limited to the situation where the n observations are a sample from a normal population. Three cases will be considered:

(A). Asymptotic efficiency of the non-parametric results as compared with the corresponding most efficient results based on the entire sample (population variance unknown).

(B). Asymptotic efficiency of the non-parametric results as compared with the corresponding most efficient results based on the $pn + O(\sqrt{n})$ smallest order statistics for the situation where the variance of the normal population is known.

(C). Asymptotic efficiency of the non-parametric results as compared with the corresponding most efficient results based on the $\beta n + O(\sqrt{n})$ smallest order statistics where β is slightly greater than p (population variance unknown).

The definition of "asymptotic" efficiency together with some of its properties is given in Section 3. Only asymptotic efficiencies will be considered.[2] However, the efficiencies obtained for the asymptotic case would seem to represent lower bounds of the efficiencies for the corresponding non-asymptotic cases since experience indicates that the efficiency of non-parametric results usually decreases as the sample size increases.

First let us consider case (A). From Theorem 3, the asymptotically most efficient results for estimating or testing the $100p\%$ population point on the basis of the entire sample (population variance unknown) are furnished by the non-central t-statistic. An expression for the asymptotic efficiency of the non-parametric results as compared with the corresponding results based on the non-central t-statistic is given in the Corollary to Theorem 3. The reciprocal of this efficiency represents the factor by which the original number of starting items must be multiplied if the non-parametric results are to asymptotically furnish the same "information" as the non-central t-statistic applied to the original number of starting items. Table 1 contains values of this factor. Although a larger number of starting items are used by the "information equivalent" non-parametric results, a noticeably smaller number of items are failed. The factor by which the number of items failed is decreased equals the value of p multiplied by the factor by which the number of starting items was increased for the "equiv-

[2] Some power function comparisons for the non-asymptotic case were given by Paul H. Jacobson in [1].

alent" non-parametric result. Table 2 contains a list of some of the resulting factors.

Next consider case (B). The first step in the analysis for this case consists in obtaining the asymptotically most efficient results. These derivations are contained in Theorems 4 and 5. The Corollary to Theorem 5 contains an expression for the asymptotic efficiency of the non-parametric results for case (B). The factor by which the original number of starting items must be multiplied to obtain "information equivalent" non-parametric results is obtained in the same way as for case (A) Table 1 lists values of this factor. In this case both the number of starting items and the number of items failed are slightly increased by use of the "equivalent" non-parametric results. The factor by which the number of items failed is increased equals the corresponding factor for the increase in number of starting items. For convenience of reference, however, values

TABLE 1

Asymptotic ratio of total numbers of items tested
(Non-parametric test over most efficient test)

Case \ p	.01	.02	.05	.10	20	.30	.40	50	.70
(A)	377%	270%	190%	160%	150%	153%	155%	157%	
(B)	101%	102%	103%	105%	109%	114%	120%	128%	164%
(C)	111%	114%	118%	122%	129%	140%	148%		

of this factor are also given in Table 2. If the variance of the normal population were unknown, the asymptotic efficiency of the non-parametric results would be at least as great as that obtained for case (B), and likely greater.

Finally consider case (C). Let p be replaced by β in Theorem 5 while the value of β corresponding to a given value of p is defined by the relation in Theorem 6. By suitable choices for the values of β and $O(\sqrt{n})$ in Theorem 5, it is possible to obtain asymptotically most efficient results for the population $100p\%$ point when the population variance is unknown and only the $\beta n + O(\sqrt{n})$ smallest values of the sample are available. These results are presented in Theorem 6. The Corollary to Theorem 6 contains an expression for the asymptotic efficiency of the non-parametric results as compared with the corresponding results of Theorem 6. The factor by which the number of starting items must be increased to obtain "equivalent" non-parametric results is computed as in cases (A) and (B). Table 1 contains values of this factor. The value of β represents the fraction of starting items which are failed if the estimates and tests of Theorem 6 are used. Table 2 contains corresponding values of β for certain values of p. The factor by which the number of items failed is decreased equals p/β times the

factor by which the number of starting items was increased to obtain the "equivalent" non-parametric results. Table 2 presents values of this factor.

The results of Theorem 6 furnish an asymptotically efficient method of estimating and testing the smaller population percentage points while only failing a small percentage of the starting items (for the case of normality). Since a larger number of items are failed and much more work is required for computing the necessary statistics, however, this method is not necessarily preferable to the non-parametric method from the viewpoint of "information" per unit cost. In many cases the difference in cost will be slight. Since the non-parametric results are valid under much more general conditions, they would seem to be preferable for these cases.

TABLE 2

Asymptotic ratio of numbers of items failed

(Non-parametric test over most efficient test)

β	.0113	0234	.0612	.130	.287	.476	.70	
Case \ p	01	02	.05	.10	20	30	.40	50
(A)	3.77%	5.40%	9.50%	16.0%	30.2%	45.9%	62.0%	78.5%
(B)	101%	102%	103%	105%	109%	114%	120%	128%
(C)	99%	98%	96%	94%	90%	88%	85%	

3. Definition of asymptotic efficiency. In this section the n observations are assumed to be a sample from a normal population Let the $100p\%$ point of the population be denoted by θ_p. Several classes of results for investigating θ_p are considered in this paper. For example, the non-parametric estimates and tests represent one class; the asymptotically most efficient results based on the entire sample (population variance unknown) represent another class; etc. The results considered consist of point estimates of θ_p, confidence intervals for θ_p, and significance tests for θ_p based on these confidence intervals. For a specified class, every point estimate and every endpoint of a confidence interval (a one-sided confidence interval has only one endpoint) consists of some statistic T whose variance is of the form $\sigma_T^2/n + o(1/n)$ for large n. Here σ_T^2 is independent of n and has the same value for all statistics T of the class. Also for every such statistic T the quantity

$$\sqrt{n}(T - \theta_p)/\sigma_T$$

has a distribution which is asymptotically normal with unit variance and some finite mean A which is independent of the unknown parameters of the normal population. By suitable choice of T, the mean A can be made to have any specified value.

Now let us define the asymptotic efficiency of the class of non-parametric results as compared to a class of results of the type defined by (A), (B) or (C). Let the non-parametric results be based on n sample values while the other class of results is based on m sample values. Let the common value of σ_T^2 for the non-parametric results be denoted by σ_1^2 while the common value of this quantity for the other class is denoted by σ_2^2. If $\sigma_1^2/n = \sigma_2^2/m$ when $m = nE$, then the asymptotic efficiency of the non-parametric results (compared to the specified class of results) is defined to be $100E\%$. For the situations considered in this paper, E is independent of n, m and the parameters of the normal population.

Asymptotic efficiency, as defined in the preceding paragraph, has the property that the statistic (or statistics) yielded by a non-parametric result based on n sample values has approximately the same distribution as the corresponding statistic (or statistics) based on m sample values from the specified class if $m = nE$ (n large). For example, consider a non-parametric unbiased estimate T_1 of θ_p based on n sample values and an unbiased estimate T_2 of θ_p from the specified class based on m sample values. Then, if $m = nE$, the distributions of

$$\sqrt{n}(T_1 - \theta_p)/\sigma_1, \qquad \sqrt{n}(T_2 - \theta_p)/\sigma_1$$

are asymptotically identical (note that $\sigma_1^2/n = \sigma_2^2/m$). Similarly for the end-points of confidence intervals Consequently the power functions of significance tests based on corresponding confidence intervals are asymptotically identical if $m = nE$. It would therefore appear that the definition chosen for asymptotic efficiency is suitable for the situations to which it is applied.

4. Notation. In this paper $t(1), \cdots, t(n)$ will represent the values of the set of all n observations arranged in increasing order of magnitude. Then

$$t(1), \cdots, t(r)$$

are the r smallest values of the set of n observations. The notation $t(r)$ has meaning only if r is an integer such that $1 \leqq r \leqq n$. Often, however, expressions of the form $t[pn + O(\sqrt{n})]$ will be encountered. In what follows, an expression of the form $t(z)$ has the interpretation t (largest integer $\leqq z$). For example,

$$t(487\tfrac{1}{2}) = t(487).$$

Also the $r = pn + O(\sqrt{n})$ smallest observations are frequently referred to, here r is interpreted to be the largest integer contained in $pn + O(\sqrt{n})$; etc.

5. Theorems and derivations. First let us consider some well known estimates and tests of the population percentage points which are based on statistics of the type used for the sign test. These estimates and tests are valid under extremely general conditions. It is not necessary that the observations be drawn from the same population or even that any two observations come from the same population. Population percentage points are not necessarily unique. The strongest continuity restriction imposed is that the population cdf be continuous at the percentage point considered. These results follow from

THEOREM 1. *Let $t(1), \cdots, t(n)$ represent the values of n observations arranged in increasing order of magnitude. The n observations are statistically independent and from populations which satisfy the conditions:*

(I). *The populations have at least one $100p\%$ point in common.*

(II) *If the populations have only one common $100p\%$ point, the cdf of each population is continuous at that point.*

Let θ_p denote the value of the common $100p\%$ point if it is unique, or the open interval of common $100p\%$ points otherwise (i.e., the interval of common $100p\%$ points with its endpoints deleted). Then asymptotically $(n \to \infty)$

(i). *$t(pn)$ is a median estimate of θ_p.*

(ii). $\Pr\{t[pn + K_\alpha\sqrt{np(1-p)}] < \theta_p\} = \Pr\{t[pn + K_\alpha\sqrt{np(1-p)}] \leqq \theta_p\} = \alpha,$

where K_α is the standardized normal deviate exceeded with probability α. Relations (i) *and* (ii) *are approximately satisfied if $pn > 5$ and $p \leqq \frac{1}{2}$.*

PROOF. This theorem is a direct application of the binomial theorem Conditions (I) and (II) assure that the equality between the probabilities in (ii) holds. Relations (i) and (ii) are obtained by using the normal approximation to the binomial theorem; this approximation is reasonably accurate if $pn > 5$ and $p \leqq \frac{1}{2}$ (see [2]).

The non-parametric confidence intervals investigated are of the forms

$$t[pn + B_1\sqrt{n} + o(\sqrt{n})] < \theta_p, \qquad t[pn + B_2\sqrt{n} + o(\sqrt{n})] > \theta_p,$$
$$t[pn + B_1\sqrt{n} + o(\sqrt{n})] < \theta_p < t[pn + B_2\sqrt{n} + o(\sqrt{n})] \qquad (B_1 < B_2),$$

(these intervals have the same confidence coefficient if $<$ is replaced by $\leqq$ and $>$ by $\geqq$). The significance tests considered are those obtained from these confidence intervals while the point estimates of θ_p are based on single order statistics of the form $t[pn + B\sqrt{n} + o(\sqrt{n})]$.

When θ_p is an open interval, (i) and (ii) need interpretation. The meaning of (i) is that the probability of $t(pn)$ exceeding every value of θ_p has the value $\frac{1}{2}$ and that the probability of it being less than all values of θ_p also has the value $\frac{1}{2}$. The inequality $t[pn + K_\alpha\sqrt{np(1-p)}] \leqq \theta_p$ has the interpretaton that every value of θ_p is greater than or equal to $t[pn + K_\alpha\sqrt{np(1-p)}]$. Similarly for $t[pn + K_\alpha\sqrt{np(1-p)}] < \theta_p$.

The purpose in introducing the case where θ_p is an open interval was to point out that situations where population percentage points are not unique cause little difficulty if suitably interpreted.

Non-parametric results of the type considered in Theorem 1 are also available when the sample size is not large. For any sample size n, if the conditions of Theorem 1 are satisfied,

$$\Pr[t(r) < \theta_p] = \Pr[t(r) \leqq \theta_p] = \sum_{s=r}^{n} \frac{n!}{s!(n-s)!} p^s(1-p)^{n-s}.$$

The probability relations in Theorem 1 were obtained by approximating this summation for large n By suitable choice of r, confidence intervals and signif-

icance tests with a wide range of satisfactory confidence coefficients and significance levels can usually be obtained for a given value of n.

The above discussion emphasizes the generality of application of the non-parametric estimates and tests. For most practical situations, however, it is permissible to assume that the observations are a random sample from a population which has a probability density function that is non-zero over the range of definition and differentiable several times. Then asymptotically $t(pn)$ is also a mean estimate of θ_p (which is now necessarily a single point). Moreover, the asymptotic distribution of $t[pn + C\sqrt{n} + o(\sqrt{n})]$ can be found in terms of p, C, θ_p and the value of the probability density function at θ_p. These results are a consequence of

THEOREM 2. *Let the population from which the n sample values were drawn have a pdf $f(t)$ such that $f(t) \neq 0$ over its range of definition and $f'(t)$ exists and is continuous in some neighborhood of $t = \theta_p$. Then the variable*

$$\sqrt{n/p(1-p)}f(\theta_p)\{t[pn + C\sqrt{n} + o(\sqrt{n})] - \theta_p\}$$

has a distribution which approaches the normal distribution with mean

$$C/\sqrt{p(1-p)}$$

and unit variance as $n \to \infty$.

PROOF. If pn is replaced by $pn + C\sqrt{n} + o(\sqrt{n})$, the method used to prove this theorem is completely analogous to the proof presented on pp. 368–69 of [3].

Now let us consider the asymptotically most efficient results for estimating and testing θ_p based on the entire set of observations for the case of a sample from a normal population (population variance unknown).

THEOREM 3. *Let the n observations be a sample from a normal population (unknown variance σ^2). Asymptotically the most efficient point estimates, confidence intervals and significance tests for θ_p using all the observations are those based on the non-central t-statistic. The value of σ_T^2 (see Section 3) for these results based on the non-central t-statistic is $\sigma^2(1 + K_p^2/2)$.*

COROLLARY. *For case (A) the asymptotic efficiency of the non-parametric results equals*

$$100(1 + K_p^2/2)/2\pi p(1-p) \exp (K_p^2) \ \%.$$

PROOF. The maximum likelihood estimate of θ_p based on all n sample values is

$$\frac{1}{n}\sum_1^n t(i) - K_p\sqrt{\sum_1^n\left[t(i) - \frac{1}{n}\sum_1^n t(j)\right]^2 \Big/ (n-1)}. \tag{1}$$

This quantity is equivalent to the non-central t-statistic, as can be seen by multiplying and dividing $[(1) - \theta_p]$ by

$$\sqrt{\sum_1^n\left[t(i) - \frac{1}{n}\sum_1^n t(j)\right]^2 \Big/ (n-1)}.$$

From maximum likelihood theory, (1) is an efficient estimate of θ_p. Asymp-

totically ($n \to \infty$) the variance of (1) is of the form

$$\sigma^2(1 + K_p^2/2)/n + o(1/n),$$

and it is easily seen that the variance of an endpoint of a confidence interval for θ_p based on the non-central t-statistic is also of this form. The corollary follows from combining Theorem 2 with Theorem 3.

Next let us investigate the situation where only the $r = pn + O(\sqrt{n})$ smallest values of a sample of size n from a normal population with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$, are available. First let us consider the asymptotic distribution of

$$\left[\frac{\sum_1^r t(i) + 2a_p(n-r)t(r)}{r + 2a_p(n-r)} - \mu + \frac{(n-r)(b_p + 2a_pK_p)}{r + 2a_p(n-r)}\sigma\right] \Big/ \frac{\sigma}{\sqrt{r + 2a_p(n-r)}}, \tag{2}$$

where

$$a_p = K_p/2\sqrt{2\pi}\,(1-p)\exp\left(\frac{1}{2}K_p^2\right) + 1/4\pi(1-p)^2\exp\left(K_p^2\right),$$

$$b_p = 1/\sqrt{2\pi}(1-p)\exp\left(\frac{1}{2}K_p^2\right).$$

This distribution is given by

THEOREM 4. *Let $t(1), \cdots, t(r)$ be the $r = pn + O(\sqrt{n})$ smallest values (arranged in increasing order of magnitude) of a sample of size n from $N(\mu, \sigma^2)$. Then asymptotically ($n \to \infty$) the distribution of (2) is $N(0, 1)$.*

COROLLARY. *Let $r = pn + C\sqrt{n} + o(\sqrt{n})$. Then as n increases the distribution of*

$$\left[\frac{\sum_1^r t(i) + 2a_p(n-r)t(r)}{r + 2a_p(n-r)} - \mu + \frac{(1-p)(b_p + 2a_pK_p)}{p + 2a_p(1-p)}\sigma\right] \Big/ \frac{\sigma}{\sqrt{r + 2a_p(n-r)}}$$

approaches the normal distribution with unit variance and mean

$$C(b_p + 2a_pK_p)/[p + 2a_p(1-p)]^{3/2}.$$

PROOF. The proof of this theorem is long and will be deferred to section 6 of the paper.

If the value of σ is known, the Corollary to Theorem 4 can be used to obtain point estimates, confidence intervals and significance tests for any population percentage point (including μ). The resulting estimates and tests are asymptotically most efficient. This follows from

THEOREM 5. *Consider the $r = pn + O(\sqrt{n})$ smallest values of a sample of size*

n from $N(\mu, \sigma^2)$ where σ^2 is known. Asymptotically $(n \to \infty)$ the variance of every unbiased estimate of μ based on only $t(1), \cdots, t(r)$ and σ^2 is greater than or equal to a quantity of the form

$$\sigma^2/n[p + 2a_p(1 - p)] + o(1/n).$$

COROLLARY. *For case (B) the asymptotic efficiency of the non-parametric results is*

$$100\left[\frac{\exp(-K_p^2)}{2\pi p(1-p)}\Bigg/\left(p + \frac{K_p \exp\left(-\frac{1}{2}K_p^2\right)}{\sqrt{2\pi}} + \frac{\exp(-K_p^2)}{2\pi(1-p)}\right)\right]\%.$$

PROOF. The proof of this theorem is similar to the proof presented for Theorem 4 and will be given in section 6 following the proof of Theorem 4.

Let p be replaced by β in Theorem 4. Even if σ is unknown asymptotically most efficient estimates and tests can be obtained for the $100p\%$ point of the population if β is defined by

$$K_p = (1 - \beta)(b_\beta + 2a_\beta K_\beta)/[\beta + 2a_\beta(1 - \beta)]. \tag{3}$$

THEOREM 6. *Let p, $(0 < p < \frac{1}{2})$, be given and β defined by (3). Let $t(1), \cdots, t(r)$ be the $r = \beta n + C\sqrt{n} + o(\sqrt{n})$ smallest values of a sample of size n from a normal population. Then asymptotically*

$$\Pr\left\{\left[\sum_1^r t(i) + 2a_\beta(n - r)t(r)\right]\Big/[r + 2a_\beta(n - r)] < \theta_p\right\}$$
$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-C(b_\beta + 2a_\beta K_\beta)/[\beta + 2a_\beta(1-\beta)]^{1/2}} e^{-x^2/2}\,dx.$$

COROLLARY. *For case (C) the asymptotic efficiency of the non-parametric results is*

$$100\left[\frac{\exp(-K_p^2)}{2\pi p(1-p)}\Bigg/\left(\beta + \frac{K_\beta \exp\left(-\frac{1}{2}K_\beta^2\right)}{\sqrt{2\pi}} + \frac{\exp(-K_\beta^2)}{2\pi(1-\beta)}\right)\right]\%.$$

PROOF. Theorem 6 is an immediate consequence of relation (3) and the Corollary to Theorem 4. The Corollary to Theorem 6 follows from Theorem 2 and Theorem 6.

6. Long proofs. This section contains the long proof of Theorem 4 and the related proof of Theorem 5.

6.1. Proof of Theorem 4. If $t(r)$ is such that

$$\mu - K_p\sigma - n^{-4/10} \leqq t(r) \leqq \mu - K_p\sigma + n^{-4/10},$$

the ratio of the value of the joint probability density function f of $t(1), \cdots, t(r)$ to the value of the function

$$\begin{aligned}\frac{n!(1-p)^{n-r}}{(n-r)!}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^r \exp\Bigg\{&-\frac{1}{2}\sum_1^r\left[\frac{t(i)-\mu}{\sigma}\right]^2 \\ &- (n-r)a\left[\frac{t(r)-\mu}{\sigma} + K_p\right]^2 - (n-r)b\left[\frac{t(r)-\mu}{\sigma} + K_p\right]\Bigg\}\end{aligned} \tag{4}$$

is of the form $1 + o(1)$. Here (and in the remainder of section 6) $a = a_p$, $b = b_p$. Also, for large n and any positive ϵ, the integral of f over the ranges of the $t(1), \cdots, t(r-1)$ and for $t(r)$ between $\mu - K_p\sigma - n^{-\frac{1}{2}+\epsilon}$ and $\mu - K_p\sigma + n^{-\frac{1}{2}+\epsilon}$ differs from unity by a quantity which is of the order $o(1)$, i.e., a quantity which $\to 0$ as $n \to \infty$.

Now consider the moment generating function of (2), i.e., $E[e^{\theta(2)}]$. In evaluating this function of θ, let the range of integration of $t(r)$, (i.e., the range after the other variables have been integrated out), be subdivided into the five intervals

$$-\infty \quad \text{to} \quad \mu - D\sigma, \qquad \mu - D\sigma \quad \text{to} \quad \mu - K_p\sigma - n^{-4/10},$$
$$\mu - K_p\sigma - n^{-4/10} \quad \text{to} \quad \mu - K_p\sigma + n^{-4/10},$$
$$\mu - K_p\sigma + n^{-4/10} \quad \text{to} \quad \mu + D\sigma, \qquad \mu + D\sigma \quad \text{to} \quad \infty.$$

Here D is a positive constant which is independent of n and such that

$$(1/D)^{n-r}(1/p)^{r-1}[1/(1-p)]^{n-r} < \exp\left[-\frac{|\theta|(n-r)(b+2aK_p)}{\sqrt{r+2a(n-r)}}\right]$$

for n sufficiently large and

$$D > |K_p| + n^{-4/10}/\sigma, \qquad 1 - N(D) = N(-D) < e^{-\frac{1}{2}D^2}/D,$$

where

$$N(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{1}{2}y^2}\,dy$$

First let us consider the interval $\mu - K_p\sigma - n^{-4/10}$ to $\mu - K_p\sigma + n^{-4/10}$. Using (4) in place of f, completing the square in the exponent, making the change of variable

$$x(i) = t(i) - \theta/\sqrt{r+2a(n-r)} \qquad (i = 1, \cdots, r),$$

integrating $x(1), \cdots, x(r-1)$ over their ranges and then $x(r)$ over the interval $\mu - K_p\sigma - n^{-4/10} - \theta/\sqrt{r+2a(n-r)}$ to

$$\mu - K_p\sigma + n^{-4/10} - \theta/\sqrt{r+2a(n-r)},$$

an expression of the form

$$\exp(\theta^2/2) + o(1) \tag{5}$$

is obtained. From the above results, this expression differs from the corresponding integration of f by a term of order $o(1)$; hence the contribution to the mgf for the interval considered is of the form (5).

Next consider the interval $\mu - K_p\sigma + n^{-4/10}$ to $\mu + D\sigma$. After $t(1), \cdots, t(r-1)$ have been integrated out, the integrand becomes

$$\begin{aligned}&\frac{n!}{(r-1)!(n-r)!}\left\{N\left[\frac{t(r)-\mu}{\sigma} - \frac{\theta}{\sqrt{r+2a(n-r)}}\right]\right\}^{r-1}\\ &\cdot\left\{1 - N\left[\frac{t(r)-\mu}{\sigma}\right]\right\}^{n-r}\exp\left\{-\frac{1}{2}\left[\frac{t(r)-\mu}{\sigma}\right]^2\right.\\ &\left.+\frac{2\theta a(n-r)}{\sqrt{r+2a(n-r)}}\left[\frac{t(r)-\mu}{\sigma} + K_p\right] + \frac{b\theta(n-r)}{\sqrt{r+2a(n-r)}}\right\}.\end{aligned} \tag{6}$$

By writing $\{N[(t(r) - \mu)/\sigma - \theta/\sqrt{r + 2a(n - r)}]\}^{r-1}$ in the form

$$\left[N\left(\frac{t(r) - \mu}{\sigma}\right)\right]^{r-1}\left\{\left[1 - \theta(1 + o(1))/\sqrt{r + 2a(n - r)}\right.\right.$$

$$\left.\left.\cdot N\left(\frac{t(r) - \mu}{\sigma}\right)\sqrt{2\pi}\exp\left(\frac{t(r) - \mu}{\sigma}\right)^2\right]^{\sqrt{n}}\right\}^{(r-1)/\sqrt{n}}$$

and maximizing $\exp\{2\theta a(n - r)[t(r) - \mu]/\sigma\sqrt{r + 2a(n - r)}\}$ with respect to $t(r)$ in the specified interval, it is seen that the value of (6) is less than an expression of the form

$$\frac{n!\exp(C_1\theta\sqrt{n})}{(r - 1)!(n - r)!}\left\{N\left[\frac{t(r) - \mu}{\sigma}\right]\right\}^{r-1}\left\{1 - N\left[\frac{t(r) - \mu}{\sigma}\right]\right\}^{n-r} + o(1)$$

for n sufficiently large. Differentiation shows that $\{N[\,]\}^{r-1}\{1 - N[\,]\}^{n-r}$ is a decreasing function of $t(r)$ in the specified interval if n is large enough. Also, if $t(r) = \mu - K_p\sigma + n^{-4/10}$, for large n the value of

$$\left\{N\left[\frac{t(r) - \mu}{\sigma}\right]\right\}^{(r-1)n^{-6/10}}$$

$$\left\{1 - N\left[\frac{t(r) - \mu}{\sigma}\right]\right\}^{(n-r)n^{-6/10}}\Big/p^{(r-1)n^{-6/10}}(1 - p)^{(n-r)^{-6/10}}$$

is less than a constant which is less than unity. Thus the value of (6) is less than a quantity of the form

$$\frac{n!\,p^{r-1}(1 - p)^{n-r}}{(r - 1)!(n - r)!}\exp(-C_3^2n^{1/10}) + o(1),$$

which in turn is less than an expression of the form

$$C_4\sqrt{n}\exp(-C_3^2n^{1/10}) + o(1)$$

for n sufficiently large. Thus the integral of (6) over the specified interval is of the order $o(1)$. An analogous proof shows that the contribution to the mgf for the interval $\mu - D\sigma$ to $\mu - K_p\sigma - n^{-4/10}$ is also of order $o(1)$.

Finally consider the interval $\mu + D\sigma$ to ∞. For large n the integral of (6) over this interval is less than an expression of the form

$$\frac{n!\,p^{r-1}(1 - p)^{n-r}}{(r - 1)!(n - r)!}\int_{\mu+D\sigma}^{\infty}\exp\left\{-\tfrac{1}{2}(n - r)\left[\frac{t(r) - \mu}{\sigma}\right]^2\right\}dt(r) + o(1); \tag{7}$$

i.e., the contribution to the mgf for this interval is of the order $o(1)$ since the coefficient of the integral is less than an expression of the form $C\sqrt{n}$. The upper limit (7) was obtained by replacing

$N\{[t(r) - \mu]/\sigma - \theta/\sqrt{r + 2a(n - r)}\}$ by 1,

$$1 - N\left[\frac{t(r) - \mu}{\sigma}\right] \text{ by } \frac{1}{D}\exp\left\{-\frac{1}{2}\left[\frac{t(r) - \mu}{\sigma}\right]^2\right\},$$

$$(1/D)^{n-r}\ (1/p)^{r-1}\,[1/(1-p)]^{n-r}\exp\,[-\theta(n-r)\ (b+2aK_p)/\sqrt{r+2a(n-r)}]$$

by 1.

A similar type proof shows that the integral of (6) from $-\infty$ to $\mu - D\sigma$ is also of the order $o(1)$.

Thus the mgf of (2) is of the form (5) for large n and Theorem 4 is verified.

6.2. Proof of Theorem 5. Let us consider a single sample value from the multivariate population consisting of the r smallest order statistics of a sample of size n from $N(\mu, \sigma^2)$, where σ^2 is known. Then the variance of every unbiased estimate of μ based on this sample and the value of σ^2 is greater than or equal to the reciprocal of

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{t(2)}\left(\frac{\partial \log f}{\partial\mu}\right)^2 f dt(1)\cdots dt(r) = -\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{t(2)}\frac{\partial^2\log f}{\partial\mu^2} f dt(1)\cdots dt(r), \tag{8}$$

where f is the joint *pdf* of the r smallest order statistics of a sample of size n from $N(\mu, \sigma^2)$. For proof of this statement see pp. 480–81 of [3]. In the lower part of (8) the variables $t(1), \cdots, t(r-1)$ can be integrated out leaving an explicit function of $t(r)$ to be integrated from $-\infty$ to ∞. To evaluate this integral for large n, choose some large but fixed interval $\mu - D\sigma$ to $\mu + D\sigma$ as was done in the proof of Theorem 4. Using a method similar to that presented on pp. 368–69 of [3], the value of the integral for the interval $\mu - D\sigma$ to $\mu + D\sigma$ is found to be of the form

$$n[p + 2a(1-p)]/\sigma^2 + o(n).$$

A procedure analogous to that used in the latter part of section 6.1 shows that integration outside this interval yields an expression of order $o(n)$.

REFERENCES

[1] Paul H Jacobson, "The relative power of three statistics," *Jour. Amer. Stat. Assn.*, Vol. 42 (1947), pp. 575–584

[2] Paul G. Hoel, *Introduction to Mathematical Statistics*, John Wiley and Sons, 1947, p. 45

[3] Harald Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.

ON THE RELATIVE EFFICIENCIES OF BAN ESTIMATES[1]

By Leo Katz

Michigan State College

1. Introduction. J. Neyman [3] defined BAN (best asymptotically normal) estimates as those functions of observed relative frequencies which i) are consistent, ii) are asymptotically normally distributed, iii) are asymptotically efficient and iv) possess continuous partial derivatives with respect to each relative frequency. He suggested the following two problems, first, to determine the class of estimates which possess the above four properties and second, to investigate this class of estimates to see whether, and under what conditions, the use of some of them is preferable to the use of others. Neyman's paper dealt with the first problem directly and with the second obliquely. With respect to the first problem, he showed that two types of χ^2-minimum estimates belong to the class of BAN estimates as do, obviously, maximum likelihood (ML) estimates. On the second problem, the χ^2-minimum estimates may be more easily computed than the corresponding ML estimates in many cases, the ease of computation being especially pronounced for the modified χ^2 with observed, rather than expected, relative frequencies in the denominators. The present paper contains some additional information regarding the relative merits of these estimates.

For simplicity, we shall consider a random variable taking on values

$$x = 0, 1, 2, 3, \cdots$$

with probabilities $p(x \mid \theta_1, \theta_2, \cdots, \theta_r)$ depending on r parameters. In working with χ^2-minimum estimates, it is almost always necessary to truncate the probability law, taking

$$\begin{aligned} f(x) &= p(x \mid \theta_1, \theta_2, \cdots, \theta_r), \qquad x = 0, 1, \cdots, k-1, \qquad \text{and} \\ f(k) &= \sum_k^\infty p(x \mid \theta_1, \theta_2, \cdots \theta_r). \end{aligned} \tag{1.1}$$

The ML estimates are asymptotically efficient, i.e., have minimum variance, with respect to the probability law, $p(x \mid \theta)$, and the χ^2 estimates have the same property with respect to the truncated p. l., $f(x \mid \theta)$. This suggests that the optimum variances of the estimates of the parameters of the two in samples of N may differ and, further, that the minimum variance of the χ^2 estimates may depend essentially upon the choice of k. In the course of some unpublished work by Evelyn Fix and others in the Statistical Laboratory at the University of California on χ^2 estimation of the parameters of several different p. l.'s the same anomalous situation occurred repeatedly. When the observed data were fitted

[1] This paper was presented to a joint meeting of the American Mathematical Society and the Institute of Mathematical Statistics at Boulder, Colorado on September 1, 1949.

by the truncated p. l. with the estimated parameters, the fit appeared to be *improved* when k was chosen smaller. This suggested that perhaps, contrary to intuition, it might be possible to improve the precision of estimation by choosing k smaller, within certain limits. This paper proves that this notion is false and that some other explanation of this phenomenon is needed.

2. Relative efficiency. Cramér [1] has shown, simultaneously with Rao [6], that under mild conditions of regularity, the variance of an unbiased estimate, $\theta^* = \theta^*(x_1, x_2, \cdots, x_N)$, of a single parameter, θ, where $x_1, x_2, \cdots, x_N$ are the observed sample, satisfies the following inequality for fixed N:

$$D^2(\theta^*) \geq \frac{1}{NE\left[\frac{\partial \log p(x)}{\partial \theta}\right]^2}, \tag{2.1}$$

the lower bound being attained only by "efficient" statistics. We may take as a measure of the relative precision attainable in the estimation of the parameter of the truncated p. l. (1.1) the ratio of the lower bounds (2.1) of variances of the estimates of the parameters of the original p. l., $p(x \mid \theta)$, and of the truncated p. l., $f(x|\theta)$. We define

$$\text{Rel. Eff.} = \frac{E\left[\frac{\partial \log f(x)}{\partial \theta}\right]^2}{E\left[\frac{\partial \log p(x)}{\partial \theta}\right]^2}. \tag{2.2}$$

In the case of functions depending on several parameters, $p(x \mid \theta_1, \theta_2, \cdots, \theta_r)$, and unbiased estimates, θ_i^*, which are functions of the observed relative frequencies, with non-singular covariance matrix $\| L_{ij} \|$, Cramér [1] showed that the fixed ellipsoid,

$$N \sum_{i=1}^{r} \sum_{j=1}^{r} \delta_{ij} t_i t_j = r + 2, \tag{2.3}$$

where

$$\delta_{ij} = E\left[\frac{\partial \log p(x)}{\partial \theta_i} \frac{\partial \log p(x)}{\partial \theta_j}\right],$$

lies wholly within the concentration ellipsoid,

$$\sum_{i=1}^{r} \sum_{j=1}^{r} L^{ij} t_i t_j = r + 2, \tag{2.4}$$

where $\| L^{ij} \| = \| L_{ij} \|^{-1}$. The two ellipsoids coincide if and only if the θ_i^* are joint efficient estimates of the θ_i. Thus, the covariance matrix of a set of joint efficient estimates is $\| N\delta_{ij} \|^{-1}$. In this case, we may define separately the relative efficiency with respect to each of the parameters as in (2.2) or we may consider the set of estimates for one function to possess greater concentration

than the set for the other function if the fixed ellipsoid (2.3) for the first lies wholly within the similar ellipsoid for the second. The latter will be the procedure we adopt in section 5.

3. Estimation of a single parameter. With $p(x \mid \theta)$ and $f(x \mid \theta)$ defined as in (1.1), form the difference

$$\phi(k) = E\left[\frac{\partial \log p(x)}{\partial \theta}\right]^2 - E\left[\frac{\partial \log f(x)}{\partial \theta}\right]^2. \tag{3.1}$$

The regularity conditions under which the Cramér-Rao inequality (2.1) holds involve existence of $\partial p(x)/\partial\theta$ for all x and absolute convergence of

$$\sum_x \frac{\partial p(x)}{\partial \theta}.$$

Assuming we have a regular case of estimation in Cramér's sense so that these conditions hold, we may write

$$\phi(k) = \sum_k^\infty \frac{1}{p(x)}\left[\frac{\partial p(x)}{\partial \theta}\right]^2 - \frac{1}{f(k)}\left[\frac{\partial f(k)}{\partial \theta}\right]^2, \tag{3.2}$$

and, since $\partial f(k)/\partial\theta = \sum_k^\infty (\partial p(x)/\partial\theta)$ by the second of the regularity conditions above and $f(k) = \sum_k^\infty p(x)$ by (1.1),

$$\phi(k)f(k) = \sum_k^\infty p(x) \sum_k^\infty \left[\frac{1}{\sqrt{p(x)}}\frac{\partial p(x)}{\partial \theta}\right]^2 - \left[\sum_k^\infty \frac{\partial p(x)}{\partial \theta}\right]^2. \tag{3.3}$$

By the Cauchy inequality, the right member of (3.3) is non-negative and, since $f(k) > 0$, it follows that $\phi(k) \geqq 0$, with the sign of equality holding only when $\partial p(x)/\partial\theta$ is proportional to $p(x)$ for all $x \geqq k$. In this event, $p(x) = K_\theta e^{g(x)}$, where K_θ is a constant depending on θ. Now, if $g(x)$ is constant, $p(x)$ is a rectangular p. l. On the other hand, if $g(x)$ is not constant, there are two cases which must be considered, namely:

$$\begin{aligned} &\text{a) } p(x) = K_\theta e^{g(x)}, && x \geqq 0, \quad \text{and} \\ &\text{b) } p(x) = p_1(x \mid \theta), && 0 \leqq x < a \leqq k, \\ &\phantom{\text{b) } p(x)} = K_\theta e^{g(x)}, && x \geqq a. \end{aligned}$$

In the first case, $K_\theta = (\sum_{x=0}^\infty e^{g(x)})^{-1}$ and is independent of θ, so that we do not have a case of estimation at all. In the second case, each $p(x)$ for $x \geqq a$ is known *a priori* to within a multiplicative constant depending on θ and, hence, no essential information is lost in truncation. Thus, except in these trivial cases, the relative efficiency is less than unity.

It then appears that, in every case of regular estimation, the variance of an efficient estimate of the parameter of the p. l. $p(x \mid \theta)$ is less than the corresponding variance for the truncated p. l. $f(x \mid \theta)$ and that, as an immediate consequence, the ML estimate in general is capable of greater precision than

the χ^2-minimum estimate for fixed N. This is the result mentioned in the first paragraph of section 1. It should be pointed out that the regularity conditions for the Cramér-Rao inequality are stringent enough to give this result. To complete the argument for estimation of a single parameter, form the function

$$\psi(k) = p(k) \sum_{k+1}^{\infty} p(x) \sum_{k}^{\infty} p(x)[\phi(k) - \phi(k+1)], \tag{3.4}$$

where $\phi(k)$ is defined by (3.1). Using (3.1) and (1.1), we may write

$$\begin{aligned}\phi(k) - \phi(k+1) = \frac{1}{p(k)}\left[\frac{\partial p(k)}{\partial\theta}\right]^2 + \frac{1}{\sum_{k+1}^{\infty} p(x)}\left[\sum_{k+1}^{\infty}\frac{\partial p(x)}{\partial\theta}\right]^2 \\ - \frac{1}{\sum_{k}^{\infty} p(x)}\left[\sum_{k}^{\infty}\frac{\partial p(x)}{\partial\theta}\right]^2.\end{aligned} \tag{3.5}$$

Making use of (3.5), straightforward algebraic reduction of (3.4) gives

$$\psi(k) = \left[\frac{\partial p(k)}{\partial\theta}\sum_{k+1}^{\infty} p(x) - p(k)\sum_{k+1}^{\infty}\frac{\partial p(x)}{\partial\theta}\right]^2 \geqq 0, \tag{3.6}$$

the sign of equality holding again only for the p. l.'s discussed after (3.3). Since the first three factors in the right member of (3.4) are positive, it follows that $\phi(k)$ is a strictly decreasing function of k. Thus, the variance of an efficient estimate of the parameter of a truncated p. l., $f(x)$, depends upon the choice of k and decreases in strictly monotone fashion to the variance of the original p. l., $p(x)$, as limit. As a result, the anomalous situation mentioned in the second paragraph of section 1 does *not* arise through irregularity in the behavior of this variance.

4. Poisson and binomial probability laws. The Poisson p. l., $p(x \mid \lambda) = e^{-\lambda}\lambda^x/x!$ gives immediately

$$E\left[\frac{\partial \log p(x)}{\partial\lambda}\right]^2 = \frac{1}{\lambda}, \tag{4.1}$$

whence, from (2.1), we obtain the usual result that the variance of the best unbiased estimate of λ is λ/N. The truncated p. l. has $\partial \log f(x)/\partial\lambda = (x/\lambda) - 1$ for $x \leqq (k-1)$, and $(\partial \log f(k))/\partial\lambda = p(k-1)/\sum_k^\infty p(x)$.
Thus,

$$E\left[\frac{\partial \log f(x)}{\partial\lambda}\right]^2 = \frac{1}{\lambda}\left[\sum_{0}^{k-1} p(x) + (\lambda - k)p(k-1)\right] + \frac{[p(k-1)]^2}{\sum_{k}^{\infty} p(x)}. \tag{4.2}$$

Writing $P(k-1)$ for $\sum_0^{k-1} p(x)$, we obtain finally,

$$\text{Rel. Eff.}_{\text{Poisson}}(k) = P(k-1) + (\lambda - k)p(k-1) + \frac{\lambda[p(k-1)]^2}{1 - P(k-1)}. \tag{4.3}$$

Values of $p(k)$ and $1 - P(k - 1)$ are given directly in Molina's Tables [2] for integer values of k and $\lambda = .001\ (.001)\ .01\ (.01)\ .3(.1)\ 15(1)\ 100$, or may be obtained indirectly from Pearson's Tables [4] of the incomplete Γ-function. In the classical example of a Poisson p. l. quoted by von Bortkiewicz, relating to numbers of deaths due to kicks by horses in Prussian Army Corps, $N = 200$ and the average number of deaths per corps-year is .61. Either χ^2 procedure would take $k = 2$ and $\lambda = .6$, approximately. Using these values, we find that Rel. Eff. $(k = 2 \mid \lambda = .6) = .9508$, i.e., the loss in efficiency incurred by using a χ^2 estimate rather than a ML estimate is of the order of five per cent.

The binomial p. l. is given by $p(x \mid n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$, $x = 0, 1, \cdots, n$, where n is a known parameter and θ is the parameter to be estimated from a sample of N observations. We obtain directly $E[(\partial \log p(x))/\partial\theta]^2 = n/(\theta(1 - \theta))$. Computing a similar quantity for the truncated p. l. and making use of the notations $p(x; n) \equiv \binom{n}{x} \theta^x (1 - \theta)^{n-x}$ and $P(a; n) \equiv \sum_0^a p(x; n)$, we obtain, after some reduction,

$$
\begin{aligned}
\text{Rel. Eff.}_{\text{binomial}}(k) = \frac{\theta}{1-\theta}\Bigg[&(n - 1)P(k - 3; n - 2) \\
&+ \frac{1 - 2n\theta}{\theta} P(k - 2; n - 1) + nP(k - 1; n) \\
&+ \frac{n\{P(k - 1; n) - P(k - 2; n - 1)\}^2}{1 - P(k - 1; n)}\Bigg].
\end{aligned} \tag{4.4}
$$

The form (4.4) is suitable for computation if tables, such as Pearson's Tables [5], of the incomplete B-function are available covering a range up to the parameter n. If such tables are not available (4.4) is inconvenient since it involves probabilities associated with three different binomial laws In this case we may use the relations

$$
\begin{aligned}
P(a; n) - P(a - 1; n - 1) &= (1 - \theta)p(a; n - 1), \\
p(a; n) &= \frac{n\theta}{a} p(a - 1; n - 1) \quad \text{and} \\
p(a; n - 1) &= \frac{(n - a)\theta}{a(1 - \theta)} p(a - 1; n - 1)
\end{aligned} \tag{4.5}
$$

to obtain the alternative form

$$
\begin{aligned}
\text{Rel. Eff.}_{\text{binomial}}(k) = {} & P(k - 1; n - 1) + (n\theta - k)p(k - 1, n - 1) \\
&+ \frac{n\theta(1 - \theta)[p(k - 1; n - 1)]^2}{1 - P(k - 1; n - 1) + \theta p(k - 1; n - 1)},
\end{aligned} \tag{4.6}
$$

which involves only the one binomial p. l., $p(x \mid n - 1, \theta)$.

As an example, consider the probability situation in which ten independent

trials are made, each with the same probability of success, θ. The number of successes in each set of ten trials is one observation. On the basis of N observations, we are to estimate θ. We shall investigate the relative efficiencies when $\theta = .10$. Taking $n = 10$ and $\theta = .10$ in (4.6) we compute the following table of relative efficiencies for different choices of k:

Relative efficiencies of χ^2 estimates in the case of the binomial p. l., $n = 10$, $\theta = .10$

k	Rel. Eff.
2	.8993
3	.9828
4	.9979
5	.9998

It is obvious from the table that the loss in efficiency is not great when $k \geqq 3$ and, hence, the variances of the χ^2 estimates are practically equal to the variance of the ML estimate. But, in ordinary practice, N, the number of sets of ten trials each, would have to be over 140 before k could be safely chosen as large as $k = 3$, and even $k = 2$ requires $N \geqq 38$. Cases in which we seek to estimate parameters on the basis of about 100 observations are not rare, in the present instance, use of a χ^2 estimate would produce about 11% greater variance than the use of a ML estimate.

The two elementary examples considered in this section provide only very fragmentary evidence of the need for caution in employing χ^2-minimum estimates; much numerical work would have to be done to provide any reliable guide to the relative efficiency of such estimates.

5. Estimation of two or more parameters. Consider the p. l. $p(x \mid \theta_1, \theta_2, \cdots, \theta_r)$, $x = 0, 1, 2, \cdots$, with ellipsoid of concentration for a set of joint efficient estimates given by (2.3). The truncated p. l. given by (1.1) has a corresponding ellipsoid of concentration

$$N \sum_{i=1}^{r} \sum_{j=1}^{r} \delta'_{ij} t_i t_j = r + 2, \tag{2.3'}$$

with $\delta'_{ij} = E\left[\dfrac{\partial \log f(x)}{\partial \theta_i} \dfrac{\partial \log f(x)}{\partial \theta_j}\right]$. We shall show, in this section, that the ellipsoid (2.3) lies wholly within (2.3′), this is so if the left member of (2.3) is uniformly greater than the left member of (2 3′), for every choice of the t_i, $i = 1, 2, \cdots, r$. Accordingly, we form the difference,

$$Q(k) = \sum_{i=1}^{r} \sum_{j=1}^{r} (\delta_{ij} - \delta'_{ij}) t_i t_j . \tag{5.1}$$

Adopting the notations,

$$p_i(x) \equiv \frac{\partial p(x)}{\partial \theta_i} \quad \text{and} \quad f_i(x) \equiv \frac{\partial f(x)}{\partial \theta_i},$$

we obtain by direct subtraction,

$$Q(k) = \sum_{i=1}^{r} \sum_{j=1}^{r} \left[\sum_{x=k}^{\infty} \frac{p_i(x)p_j(x)}{p(x)} - \frac{f_i(k)f_j(k)}{f(k)} \right] t_i t_j . \tag{5.2}$$

Equation (5.2) is unchanged if the right member is written in the form

$$\begin{aligned} Q(k) = \sum_{i=1}^{r} \sum_{j=1}^{r} \Bigg[\sum_{x=k}^{\infty} \Bigg\{ & \frac{p_i(x)p_j(x)}{p(x)} - \frac{f_i(k)}{f(k)} p_j(x) \\ & - \frac{f_j(k)}{f(k)} p_i(x) + \frac{f_i(k)f_j(k)}{f(k)} \frac{p(x)}{f(k)} \Bigg\} \Bigg] t_i t_j . \end{aligned} \tag{5.3}$$

If this latter is now written as

$$Q(k) = \sum_{i=1}^{r} \sum_{j=1}^{r} f(k) \left[\frac{1}{f(k)} \sum_{x=k}^{\infty} \left\{ \left(\frac{p_i(x)}{p(x)} - \frac{f_i(k)}{f(k)} \right) \left(\frac{p_j(x)}{p(x)} - \frac{f_j(k)}{f(k)} \right\} p(x) \right] t_i t_j , \tag{5.4}$$

it is evident that the expression in square brackets in the right hand member is precisely the mean value of the expression in curly brackets taken over the set $x \geqq k$. If we denote by $\underset{x \geqq k}{E} \{g(x)\}$ the expected value of $g(x)$ over the set $x \geqq k$, we have

$$Q(k) = \sum_{i=1}^{r} \sum_{j=1}^{r} f(k) \underset{x \geqq k}{E} \left\{ \left(\frac{p_i(x)}{p(x)} - \frac{f_i(k)}{f(k)} \right) t_i \left(\frac{p_j(x)}{p(x)} - \frac{f_j(k)}{f(k)} \right) t_j \right\}. \tag{5.5}$$

Finally, since the (finite) sum of the expected values is equal to the expected value of the sum, we have,

$$Q(k) = f(k) \underset{x \geqq k}{E} \left\{ \sum_{i=1}^{r} \left[\frac{p_i(x)}{p(x)} - \frac{f_1(k)}{f(k)} \right] t_i \right\}^2 . \tag{5.6}$$

Since $f(k) > 0$, $Q(k) \geqq 0$. We need only note that $Q(k) = 0$ only if the linear form in curly brackets in (5.6) is identically zero, i.e., if each coefficient of t_i vanishes. This can happen only in the trivial cases analogous to those described in Section 3.

It has been shown that the ellipsoid of concentration of a set of joint efficient estimates of the parameters of a p. l. lies wholly within the corresponding ellipsoid of the truncated p. l. Therefore, the best procedure for estimating the parameters of a truncated p. l. cannot attain the precision of an efficient procedure for estimating those of the original p. l.

In order to complete the argument for the general case, we form the difference

$$\begin{aligned} Q(k) - Q(k+1) = \sum_{i=1}^{r} \sum_{j=1}^{r} \Bigg[& \frac{p_i(k)p_j(k)}{p(k)} - \frac{f_i(k)f_j(k)}{f(k)} \\ & + \frac{f_i(k+1)f_j(k+1)}{f(k+1)} \Bigg] t_i t_j . \end{aligned} \tag{5.7}$$

Making use of the two relationships $f(k) = p(k) + f(k+1)$ and $f_i(k) =$

$p_i(k) + f_i(k + 1)$, we have

$$(5.8) \qquad Q(k) - Q(k+1) = \frac{p(k)f(k+1)}{f(k)} \left\{ \sum_{i=1}^{r} \left[\frac{p_i(k)}{p(k)} - \frac{f_i(k+1)}{f(k+1)} \right] t_i \right\}^2.$$

The right member of (5.8) being positive except in the trivial cases, it is clear that $Q(k)$ is a strictly monotone function of k.

6. Conclusions. It has been shown that the efficiency of χ^2-minimum estimates, or any other estimates which involve computation in terms of a truncated p. l., is necessarily less than the efficiency of corresponding ML or other estimates based on the original p. l. and, further, that the efficiency increases with the point of truncation. This was established for estimates of a single parameter and, also, for joint estimates of several parameters. Examples given indicate that, in any case of regular estimation, use of χ^2-minimum estimates rather than ML estimates should be accompanied by an investigation into the loss in efficiency.

The author is indebted to Professor J. Neyman, who suggested the problem.

REFERENCES

[1] H. Cramér, "Contributions to the theory of statistical estimation," *Skandinavisk Aktuarietidskrift*, Vol. 29 (1946), pp. 85–94.

[2] E. C. Molina, *Poisson's Exponential Binomial Limit*, Van Nostrand, 1942

[3] J. Neyman, "Contribution to the theory of the χ^2 test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1949, pp. 239–273.

[4] K. Pearson, *Tables of the Incomplete* Γ*-function*, Cambridge University Press, 1922.

[5] K. Pearson, *Tables of the Incomplete* B*-function*, Cambridge University Press, 1934.

[6] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math Soc.*, Vol. 37 (1945), pp 81–91.

UNBIASED ESTIMATES WITH MINIMUM VARIANCE

By Charles Stein

University of California, Berkeley

Summary. Subject to certain restrictions, a characterization of unbiased estimates with minimum variance is obtained. For two fairly broad classes of problems, solutions are given which are more readily applicable. These are used to obtain such estimates in some particular cases. The applicability of the results to problems of sequential estimation is pointed out. The problem of unbiased estimation is not at present of much practical importance, but is of some theoretical interest and has been treated by many statisticians. Also, the method used in this paper may be applicable to other problems in statistics.

1. Introduction. Let R be a space of points x, B an additive class of subsets C of R and μ a measure over B such that R can be represented as the union of a countable collection of elements of B each of which has finite μ-measure. Let Ω be a set called the parameter space and let X be a random variable distributed in accordance with the probability density function $p(x \mid \theta)$ for some $\theta \in \Omega$, so that for any $C \in B$

$$P\{X \in C \mid \theta\} = \int_C p(x \mid \theta)\, d\mu(x).$$

A measurable real-valued function $f(x)$ on R is called an unbiased estimate of the real-valued function $g(\theta)$ on Ω if, for every $\theta \in \Omega$

$$E(f(X) \mid \theta) = \int f(x)p(x \mid \theta)\, d\mu(x) = g(\theta). \tag{1}$$

The problem considered in this paper is that of finding an unbiased estimate f^* of g which minimizes the variance at θ_0. Since this variance is

$$\begin{aligned} &E([f(X) - g(\theta_0)]^2 \mid \theta_0) \\ &\qquad = \int [f(x) - g(\theta_0)]^2 p(x \mid \theta)\, d\mu(x) \\ &\qquad = \int [f(x)]^2 p(x \mid \theta)\, d\mu(x) - \left[\int f(x)p(x \mid \theta_0)\, d\mu(x)\right]^2, \end{aligned} \tag{2}$$

this problem is equivalent to minimizing

$$\int [f(x)]^2 p(x \mid \theta_0)\, d\mu(x) \tag{3}$$

subject to (1). It will be convenient to introduce the measure

$$\nu(C) = \int_C p(x \mid \theta_0)\, d\mu(x) \tag{4}$$

and the probability ratios

$$\pi(x \mid \theta) = \frac{p(x \mid \theta)}{p(x \mid \theta_0)}. \tag{5}$$

We suppose $\pi(x \mid \theta)$ finite for almost all x, and all θ. When we say "for almost all x," we mean "except for a set of μ-measure 0."

In most practical problems, the set R is a subset of some finite-dimensional Euclidean space and μ is either ordinary Lebesgue measure or, in the case where R is countable, counting measure which makes the measure of a set the number of points it contains. An exception is the application to sequential analysis considered in section 3 below, in which R is a countable union of sets, each of which is a subset of a finite dimensional Euclidean space. For the basic notation and concepts of the theory of integration see Saks [2], Ch. I.

We shall define

$$A(\theta_1, \theta_2) = \int \pi(x \mid \theta_1)\pi(x \mid \theta_2)\, d\nu(x), \tag{6}$$

and suppose

$$A(\theta, \theta) < \infty \text{ for all } \theta. \tag{7}$$

By Schwartz's inequality this implies that $A(\theta_1, \theta_2) < \infty$ for all θ_1, θ_2. If (7) is not true then it may happen that there exists no unbiased estimate with minimum variance even though there exist unbiased estimates. Consider, for example, the case where Ω consists of two point, 0 and 1, and $g(\theta) = \theta$, and

$$p(x \mid 0) = \begin{cases} 1 \text{ for } 0 < x < 1 \\ 0 \text{ otherwise} \end{cases}$$

$$p(x \mid 1) = \begin{cases} \frac{1}{2}x^{-\frac{1}{2}} \text{ for } 0 < x < 1 \\ 0 \text{ otherwise} \end{cases}$$

and μ is ordinary Lebesgue measure. It is clear that there exist unbiased estimates of θ with arbitrarily small positive variance at $\theta = 0$ but there exists none with 0 variance.

2. The principal theorem. In accordance with the usual terminology we denote by L_2 the class of all measurable functions ϕ such that

$$\int [\phi(x)]^2\, d\nu(x) < \infty. \tag{8}$$

Finally, G is the class of all functions ψ expressible in the form

$$\psi(\theta) = \int \phi(x)\pi(x \mid \theta)\, d\nu(x) \quad \text{with} \quad \phi \in L_2. \tag{9}$$

THEOREM 1. *If $\pi(x \mid \theta)$ is finite for all θ and almost all x, and (7) is satisfied, and there exists an unbiased estimate of g, then there exists an unbiased estimate*

f^* of g which minimizes (3). *If f^* has finite variance then any other unbiased estimate of g with minimum variance at θ_0 is essentially equal to f^*, that is, differs from f^* only on a set of μ-measure* 0. *A function f is an unbiased estimate of g with minimum variance at θ_0 if and only if there exists a real-valued functional T on G for which*

$$TA(\theta, \theta_1) = g(\theta_1) \text{ for all } \theta_1 \in \Omega, \tag{10}$$

$$T \int \phi(x)\pi(x \mid \theta)\, d\nu(x) = \int \phi(x) f(x)\, d\nu(x) \text{ for all } \phi \in L_2. \tag{11}$$

(*The preceding sentence does not assume the existence of an unbiased estimate of g.*) *The minimum variance is $Tg(\theta) - [g(\theta_0)]^2$.*

Proof. Let $\{f_i\}$ be a sequence of unbiased estimates of g such that

$$\lim_{i\to\infty} \int [f_i(x)]^2\, d\nu(x) = \operatorname*{g.l.b.}_f \int [f(x)]^2\, d\nu(x)$$

where f ranges over all unbiased estimates of g. Then by the weak compactness of every sphere in L_2 (see [1], p. 10) there exists $f^* \in L_2$ and an increasing sequence $\{n_i\}$ of integers for which

$$\int \phi f^*\, d\nu = \lim_{i\to\infty} \int \phi f_{n_i}\, d\nu \text{ for all } \phi \in L_2.$$

Since $\pi(x \mid \theta) \in L_2$ by (7), this implies that f^* is an unbiased estimate of g. Also

$$\int [f^*]^2\, d\nu \le \lim_{i\to\infty} \int f_{n_i}^2\, d\nu = \operatorname*{g.l.b.}_f \int f^2\, d\nu. \tag{12}$$

Thus f^* is an unbiased estimate of g with minimum variance.

Let $\phi_1 \in L_2$ be such that

$$\int \phi_1(x)\pi(x \mid \theta)\, d\nu(x) = 0 \text{ for all } \theta \in \Omega. \tag{13}$$

Then, using the f^* defined in the last paragraph, we obtain for any real ε

$$0 \le \int (f^* + \varepsilon\phi_1)^2\, d\nu - \int [f^*]^2\, d\nu = 2\varepsilon \int \phi_1 f^*\, d\nu + \varepsilon^2 \int \phi_1^2\, d\nu \tag{14}$$

since $f^* + \varepsilon\phi_1$ is an unbiased estimate of g. Dividing (14) by ε and letting $\varepsilon \to 0$ we obtain

$$\int \varphi_1 f^*\, d\nu = 0. \tag{15}$$

If a function in G can be represented in two ways,

$$\int \phi(x)\pi(x \mid \theta)\, d\nu(x) = \int \varphi'(x)\pi(x \mid \theta)\, d\nu(x),$$

and consequently $\phi_1 = \phi' - \phi$ satisfies (13) and (15) Thus (11) defines a functional on G in a consistent way Also, this functional satisfies (10) since

$$TA(\theta, \theta_1) = T \int \pi(x \mid \theta_1)\pi(x \mid \theta)\, d\nu(x)$$

$$= \int \pi(x \mid \theta_1) f^*(x)\, d\nu(x) = g(\theta_1).$$

By (2) and (11) the minimum variance is

$$\int [f^*(x)]^2\, d\nu(x) - [g(\theta_0)]^2 = T \int f^*(x)\pi(x \mid \theta)\, d\nu(x) - [g(\theta_0)]^2$$

$$= Tg(\theta) - [g(\theta_0)]^2.$$

To prove the converse, let f^* be any function in L_2 for which there exists a functional T satisfying (10) and (11) By (11) with $\phi(x) = \pi(x \mid \theta_1)$,

$$\int f^*(x)\pi(x \mid \theta_1)\, d\nu(x) = T \int \pi(x \mid \theta)\pi(x \mid \theta_1)\, d\nu(x)$$

$$= TA(\theta, \theta_1) = g(\theta_1)$$

by (10), so that f^* is an unbiased estimate of g. Any other unbiased estimate f of g with finite variance at θ_0 is an element of L_2 Thus from (1) and (11) we obtain

$$Tg(\theta) = \int ff^*\, d\nu$$

$$= \int [f^*]^2\, d\nu.$$

Applying Schwartz's inequality to the middle expression we obtain

$$\int [f^*]^2\, d\nu \leq \int [f]^2\, d\nu$$

with strict inequality unless f is essentially equal to f^*.

COROLLARY 1. *Suppose $\pi(x \mid \theta)$ is finite for all θ and almost all x and (7) holds. Let $H_1(x, d)$ be the set of all $\theta \in \Omega$ such that $\pi(x \mid \theta) > d$, and let H be the smallest additive class containing all $H_1(x, d)$. Suppose there exists an additive set function λ over H such that there exists a finite collection of parameter points θ_k and positive number c_k such that*

$$\int \pi(x \mid \theta) \mid d\lambda(\theta) \mid \leq \sum c_k\, \pi(x \mid \theta_k) \tag{16}$$

for almost all x, and

$$\int A(\theta, \theta_1)\, d\lambda(\theta) = g(\theta_1). \tag{17}$$

Then the unbiased estimate of $g(\theta)$ with minimum variance at θ_0 is

$$f^*(x) = \int \pi(x \mid \theta)\, d\lambda(\theta) \tag{18}$$

and the minimum variance is

$$\int g(\theta)\, d\lambda(\theta) - [g(\theta_0)]^2. \tag{19}$$

Proof: We need only show that (10) and (11) are satisfied by

$$T\psi(\theta) = \int \psi(\theta)\, d\lambda(\theta)$$

and (18). But

$$TA(\theta, \theta_1) = \int A(\theta, \theta_1)\, d\lambda(\theta) = g(\theta_1) \tag{20}$$

by (17) and

$$\begin{aligned} T \int \phi(x)\pi(x \mid \theta)\, d\nu(x) &= \int d\lambda(\theta) \int \phi(x)\pi(x \mid \theta)\, d\nu(x) \\ &= \int \phi(x)\, d\nu(x) \int \pi(x \mid \theta)\, d\lambda(\theta) = \int \phi(x) f^*(x)\, d\nu(x). \end{aligned} \tag{21}$$

Since each of the functions $\phi(x)$, $\pi(x \mid \theta)$ considered as a function of x and θ is measurable (BH), their product is also. The interchange of order of integration in (21) is justified by Fubini's Theorem (Saks [2], p. 87) and (16) which by (9) implies that $\int | d\lambda(\theta) | \int \phi(x)\pi(x \mid \theta)\, d\nu(x) < \infty$. The equations (20) and (21) are equivalent to (10) and (11) respectively.

Corollary 2. *Suppose $\pi(x \mid \theta)$ is finite for all θ and almost all x and* (7) *holds. Suppose also that Ω is a set of real numbers and:*

(i) *for some m, either a positive integer or $+\infty$, $\pi(x \mid \theta)$ is, for almost all x, differentiable m times with respect to θ at $\theta = \theta_0$,*

(ii) *for each $n < m$ there exists a finite collection of parameter values $\theta_{n,k}$ and positive constants $c_{n,k}$ such that*

$$\left| \frac{\pi^{(n)}(x \mid \theta_0 + \delta) - \pi^{(n)}(x \mid \theta_0)}{\delta} \right| \leq \sum_k c_{n,k}\, \pi(x \mid \theta_{n,k}) \tag{22}$$

for all δ whose absolute value is sufficiently small and almost all x,

(iii) *there exist constants a_n such that for all θ_1,*

$$g(\theta_1) = \sum_{n=0}^{m} a_n \left[\frac{\partial^n}{\partial \theta^n} A(\theta, \theta_1) \right]_{\theta=\theta_0}, \tag{23}$$

(iv) *there exists a finite collection of parameter values θ_k and positive constants*

c_k such that

$$\sum_{n=0}^{m} \left| a_n \left[\frac{\partial^n}{\partial \theta^n} \pi(x \mid \theta) \right]_{\theta=\theta_0} \right| \le \sum_k c_k \pi(x \mid \theta_k). \tag{24}$$

Then the unbiased estimate of $g(\theta)$ with minimum variance at θ_0 is

$$f^*(x) = \sum_{n=0}^{m} a_n \left[\frac{\partial^n}{\partial \theta^n} \pi(x \mid \theta) \right]_{\theta=\theta_0}. \tag{25}$$

The minimum variance is

$$\sum_{n=0}^{m} a_n \left[\frac{\partial^n}{\partial \theta^n} g(\theta) \right]_{\theta=\theta_0}.$$

PROOF. We need only show that the functional T defined by

$$T \int \phi(x)\pi(x \mid \theta)\, d\nu(x) = \sum_{n=0}^{m} a_n \frac{\partial^n}{\partial \theta^n} \int \phi(x)\pi(x \mid \theta)\, d\nu(x) \Big]_{\theta=\theta_0} \tag{26}$$

satisfies (10) and (11) with f^* given by (25). Equation (23) yields (11) immediately. Also

$$T \int \phi(x)\pi(x \mid \theta)\, d\nu(x) = \sum_{n=0}^{m} a_n \frac{\partial^n}{\partial \theta^n} \int \phi(x)\pi(x \mid \theta)\, d\nu(x) \Big]_{\theta=\theta_0}$$

$$= \Sigma\, a_n \int \phi(x) \frac{\partial^n}{\partial \theta^n} \pi(x \mid \theta) \Big]_{\theta=\theta_0} d\nu(x)$$

by (9), (i), (22) and Lebesgue's Theorem on term by term integration (Saks [2] p. 29.). Using (24) and Lebesgue's Theorem, we find that this is equal to

$$\int \phi(x) \Sigma\, a_n \frac{\partial^n}{\partial \theta^n} \pi(x \mid \theta) \Big]_{\theta=\theta_0} d\nu(x) = \int \phi(x) f^*(x)\, d\nu(x).$$

which completes the proof.

There is an obvious combination of Corollaries 1 and 2 which will not be stated explicitly. Also Corollary 2 can be extended to involve differentiation with respect to several parameters. It would be of considerable interest to obtain a characterization of all possible functionals T in terms of the usual operations such as integration and differentiation. Also, the methods used here should be applicable, with some modifications, to other problems of minimization subject to an infinite set of side conditions.

COROLLARY 3. *Suppose that subject to the condition of Theorem 1, for $i = 1, 2, f_i^*$ are unbiased estimates of g_i with minimum variance at θ_0. Then $f_1^* + f_2^*$ is an unbiased estimate of $g_1 + g_2$ with minimum variance at θ_0.*

This follows immediately from (11) and (12) in Theorem 1. Actually, the restriction to problems satisfying the conditions of Theorem 1 is unnecessary, but we shall not prove this here.

3. Some special cases. We first consider a problem which is of little practical interest but serves well as an illustration of Corollary 1. Let X be a single observation from a uniform distribution on the interval $(\theta, \theta + 1)$, i.e.

$$p(x \mid \theta) = \begin{cases} 1 \text{ if } \theta < x < \theta + 1 \\ 0 \text{ otherwise.} \end{cases}$$

We suppose θ lies in the interval $(-N, N - 1)$ where N is a given positive integer, and take as the distribution for which the variance is to be minimized

$$p(x \mid \theta_0) = \begin{cases} \frac{1}{2N} \text{ if } -N < x < N \\ 0 \text{ otherwise.} \end{cases}$$

This is the same as using the original p.d.f. $p(x \mid \theta)$ with θ a random variable taking on the values $-N, -N + 1, \cdots, N - 1$ with equal probability. The measure μ is of course ordinary Lebesgue measure. Then

$$\pi(x \mid \theta) = \begin{cases} 2N \text{ if } \theta < x < \theta + 1 \\ 0 \text{ otherwise} \end{cases} \tag{27}$$

and

$$\frac{1}{2N} A(\theta_1, \theta_2) = \begin{cases} 0 & \text{if } \theta_1 < \theta_2 - 1 \\ \theta_1 - \theta_2 + 1 & \text{if } \theta_2 - 1 < \theta_1 < \theta_2 \\ \theta_2 - \theta_1 + 1 & \text{if } \theta_2 < \theta_1 < \theta_2 + 1 \\ 0 & \text{if } \theta_2 + 1 < \theta_1 . \end{cases} \tag{28}$$

For $-N < \theta_1 < N - 1$, equation (17) becomes

$$\int_{\max(-N, \theta_1 - 1)}^{\theta} (\theta - \theta_1 + 1)\, d\lambda(\theta) + \int_{\theta_1}^{\min(N-1, \theta_1+1)} (\theta_1 - \theta + 1)\, d\lambda(\theta) = g(\theta_1)/2N \tag{29}$$

and (18) becomes

$$f^*(x)/2N = \lambda(\min[N - 1, x]) - \lambda(\max[-N, x - 1]). \tag{30}$$

The reader will not be confused by the use of λ as a point function here, and as a set function in Corollary 1. Using (30) and integration by parts (Saks [2], p. 102) we can rewrite (29) as

$$\int_{\theta_1}^{\theta_1+1} f^*(x)\, dx = g(\theta_1), \tag{31}$$

which is merely the condition that f^* be an unbiased estimate of g. It is clear from (31) that g admits an unbiased estimate if and only if it is absolutely

continuous. Differentiating (31) we obtain

$$f^*(\theta + 1) - f^*(\theta) = g(\theta). \tag{32}$$

Consequently the general solution of (31) is

$$f^*(\theta) = \sum_{i=1}^{[\theta]+N} g'(\theta - i) + \gamma(\theta), \tag{33}$$

where γ is a function of period 1 such that

$$\int_0^1 \gamma(\theta)\, d\theta = 0. \tag{34}$$

Here, contrary to the usual convention, $[\theta]$ denotes the largest integer less than θ. The one of (33) which minimizes the variance at θ_0 is determined by the condition that there exist λ satisfying (30). Let y be any number on the half-closed interval $(-N, -N + 1)$, and sum (30) for $x = y, y + 1 \cdots y + 2N - 1$. This yields

$$\frac{1}{2N} \sum_{j=0}^{2N-1} f^*(y + j) = \lambda(N - 1) - \lambda(-N). \tag{35}$$

Carrying out the same computation on (33) we obtain

$$\frac{1}{2N} \sum_{j=0}^{2N-1} \sum_{i=1}^{j+N} g'(y + j - i) + \gamma(y) = \lambda(N - 1) - \lambda(-N). \tag{36}$$

Combining (34) and (35) we find that the proper choice of γ is that which gives

$$\begin{aligned} f^*(x) = &\sum_{i=0}^{[x]+N+1} \frac{1 + i}{2N} g'(x - [x] - N + i) \\ &+ \sum_{i=x+N}^{2N-2} \left(\frac{1 + i}{2N} - 1\right) g'(x - [x] - N + i) \\ &+ \frac{1}{2N} \sum_{j=1}^{2N-1} \{g(j - N) - g(-N)\}. \end{aligned} \tag{37}$$

If the limit of (37) as $N \to \infty$ exists, it agrees with Norlund's simplest definitions of the principal solution of (32) (see Milne-Thompson [3] formula (2) p. 201) whenever the latter is applicable. The author has not checked the agreement with Nörlund's more general definitions.

Next we consider the problem of obtaining an unbiased estimate of $g(\theta)$ with minimum variance at θ_0 when X consists of n independent observations, each uniformly distributed over the interval $(0, \theta)$. Here θ is an unknown positive number. The result is independent of the choice of θ_0. Clearly a necessary and sufficient condition for the existence of an unbiased estimate of g is that g be absolutely continuous Corollary 1 can be applied to obtain as the best unbiased estimate $g(Y) + \frac{Y}{n} g'(Y)$ where $Y = \max(X_1 \cdots X_n)$. However, this result can be obtained much more simply by observing that, given any sufficient statistic Z,

there exists an unbiased estimate with minimum variance which is a function only of Z. A proof of this is given by Blackwell [4]. But Y is a sufficient statistic, and the condition that $f^*(Y)$ be an unbiased estimate of g is that

$$\frac{n}{\theta^n} \int_0^\theta f^*(y) y^{n-1}\, dy = g(\theta).$$

This has as its unique solution that given above.

A similar situation holds when the X_i, $i = 1 \cdots n$, are independently normally distributed with unknown common mean θ and unit variance. Here Corollary 1 is not applicable, but Corollary 2 is. The result can again be obtained more simply as the unique solution of the integral equation

$$\frac{1}{\sqrt{2\pi}} \int f_0^*(y)^{-\frac{1}{2}(y-\theta\sqrt{n})^2}\, dy = g(\theta)$$

with

$$f^*(x_1, \cdots, x_n) = f_0^*(y), \qquad y = \frac{1}{\sqrt{n}} \sum_1^n x_i .$$

It should be observed that the methods of section 2 are applicable also to problems of sequential estimation. Let $X_1, X_2, \cdots$ be a sequence of real-valued random variables such that $(X_1, \cdots, X_n)$ have the joint p.d.f. $p_n(x_1, \cdots, x_n \mid \theta)$ for some unknown $\theta \in \Omega$. Suppose it has been decided to terminate the procedure on the m^{th} observation if $(X_1, \cdots, X_m) \in \bar{R}_m$ for some given sets $\bar{R}_m$ in m space, and suppose these sets are so chosen that the probability of termination is 1 for all θ. Then we can define the space $R = \bigcup_m \bar{R}_m$, the union of the $\bar{R}_m$, the measure

$$\mu(A) = \sum_m \mu_m(AR_m)$$

for any set $A \subset R$ for which the intersections $A \cap R_m$ are Borel sets, where μ_m is ordinary m-dimensional Lebesgue measure, and the probability density functions

$$p(x \mid \theta) = p_m(x_1 \cdots x_m \mid \theta) \quad \text{if} \quad x = (x_1 \cdots x_m) \in \bar{R}_m .$$

The previous results are then applicable. Most of the familiar results in the theory of statistical inference can be extended to sequential problems in the same way. Of course the interesting and difficult problems of sequential analysis are usually concerned chiefly with the appropriate choice of the regions $\bar{R}_m$.

4. Connections with the work of other authors. Many lower bounds for the variance of an unbiased estimate were obtained by Bhattacharyya [5], and some results were obtained earlier by others whose results are referred to by Bhattacharyya. His work has been extended to sequential problems as indicated in section 3 above by G. R. Seth in a doctoral dissertation at Columbia University. This leads to results analogous to, but in some respects more general than those of Wolfowitz [6]. Among other papers on sequential estimation,

there are the one by Blackwell [4] already referred to, and the one by Girshick, Mosteller, and Savage [7]. These deal mainly with problems in which there is a unique unbiased estimate based on a sufficient statistic.

The author is indebted to A. Wald, J L. Hodges, E. Barankin, and H. Rubin for some helpful suggestions and comments.

REFERENCES

[1] B. v. Sz. Nagy, "Spektraldarstellung linearer Transformationen des Hilbertschen Raumes," *Ergebnisse der Mathematik*, Vol. 5, No. 5 (1942)
[2] S. Saks, *Theory of the Integral*, Monografie Matematyczne, Tom VII, Warsaw, 1937.
[3] L. M. Milne-Thompson, *The Calculus of Finite Differences*. Macmillan, London, 1933
[4] D. Blackwell, "Conditional expectation and unbiased sequential estimation," *Annals of Math. Stat.*, Vol 18 (1947), p 105
[5] A Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation," *Sankhya*, Vol. 8 (1946), p 1 and Vol. 8 (1947), p 201
[6] J Wolfowitz, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Annals of Math Stat*, Vol 18 (1947), p. 215
[7] M. Girshick, F. Mosteller, and L Savage, "Unbiased estimates for certain binomial sampling problems," *Annals of Math. Stat*, Vol 17 (1946), p 13

DISTRIBUTION OF MAXIMUM AND MINIMUM FREQUENCIES IN A SAMPLE DRAWN FROM A MULTINOMIAL DISTRIBUTION

By Robert E. Greenwood and Mark O. Glasgow

University of Texas

1. Introduction. In this paper, the expected values

$$E\left[\begin{matrix}\max\\ \min\end{matrix}(n_1, n_2, \cdots, n_l)\right] \tag{1.1}$$
$$= \sum_{n_1+n_2+\cdots n_k=N} \frac{N!}{n_1!n_2!\cdots n_k!}\left[\begin{matrix}\max\\ \min\end{matrix}(n_1, n_2, \cdots, n_l)\right] p_1^{n_1}p_2^{n_2}\cdots p_k^{n_k}$$

will be studied The quantities $\{n_i\}$, $i = 1, 2, \cdots, k$, are understood to be non-negative integers, and the quantities $\{p_i\}$ are non-negative probabilities, $\Sigma p_i = 1$. Also, $l \leqq k$. Form (1.1) will be evaluated for the binomial case $l = k = 2$ and for the special trinomial case $p_1 = p_2$ with $l = 2$, $k = 3$.

2. Binomial distribution. The evaluations for the expected values in the binomial case can be given explicitly in terms of the incomplete Beta function. This function may be defined by the relation

$$I_q(n - k, k + 1) = \sum_{r=0}^{k} \binom{n}{r} (1 - q)^r q^{n-r}, \tag{2.1}$$

whence

$$I_{1-q}(k + 1, n - k) = \sum_{r=k+1}^{n} \binom{n}{r} (1 - q)^r q^{n-r}.$$

It is seen that

$$I_q(n - k, k + 1) + I_{1-q}(k + 1, n - k) = 1. \tag{2.2}$$

For the binomial case, $n_2 = N - n_1$ and $p_2 = 1 - p_1$, and thus instead of (n_1, n_2) and (p_1, p_2) one may use $(n, N - n)$ and $(p, 1 - p)$ without any subscripts and without sacrifice of clarity. This will be done in some instances in what follows. The evaluation of

$$E\left[\begin{matrix}\max\\ \min\end{matrix}(n_1, n_2)\right] = \sum_{n=0}^{N} \binom{N}{n}\left[\begin{matrix}\max\\ \min\end{matrix}(n, N - n)\right] p^n(1 - p)^{N-n} \tag{2.3}$$

is slightly different for the two cases N odd and N even.

For N odd, and for the minimum form, the summation may be written in two parts, (a) and (b),

$$0 \leqq n \leqq \frac{N - 1}{2}, \tag{a}$$

in which range $\min(n, N - n) = n$, and

(b) $$\frac{N+1}{2} \leqq n \leqq N,$$

in which range $\min(n, N - n) = N - n$. In the (a) part summation one gets

$$\sum_{n=0}^{(N-1)/2} \binom{N}{n} np^n(1-p)^{N-n} = \sum_{n=1}^{(N-1)/2} \binom{N-1}{n-1} Npp^{n-1}(1-p)^{N-n}$$

$$= Np \sum_{r=0}^{(N-3)/2} \binom{N-1}{r} p^r(1-p)^{N-1-r} = NpI_{1-p}\left(\frac{N+1}{2}, \frac{N-1}{2}\right).$$

In the (b) part summation one gets

$$\sum_{n=(N+1)/2}^{N} \binom{N}{n} (N-n)p^n(1-p)^{N-n} = \sum_{(N+1)/2}^{N-1} \binom{N-1}{n}$$

$$\cdot N(1-p)p^n(1-p)^{N-n-1} = N(1-p)I_p\left(\frac{N+1}{2}, \frac{N-1}{2}\right).$$

Similar algebraic manipulations, supplemented by symmetry, can be used to effect the evaluations tabulated below.

For N odd there result the forms

$$E[\min(n_1, n_2)] = NpI_{1-p}\left(\frac{N+1}{2}, \frac{N-1}{2}\right)$$

$$+ N(1-p)I_p\left(\frac{N+1}{2}, \frac{N-1}{2}\right),$$

(2.4)

$$E[\max(n_1, n_2)] = NpI_p\left(\frac{N-1}{2}, \frac{N+1}{2}\right)$$

$$+ N(1-p)I_{1-p}\left(\frac{N-1}{2}, \frac{N+1}{2}\right).$$

For N even there result the forms

(2.5)

$$E[\min(n_1, n_2)] = NpI_{1-p}\left(\frac{N}{2}, \frac{N}{2}\right) + N(1-p)I_p\left(\frac{N}{2}+1, \frac{N}{2}-1\right),$$

$$E[\max(n_1, n_2)] = NpI_p\left(\frac{N}{2}, \frac{N}{2}\right) + N(1-p)I_{1-p}\left(\frac{N}{2}-1, \frac{N}{2}+1\right).$$

For this simple binomial case, $\max(n_1, n_2) + \min(n_1, n_2) = N$ and linearity in the expected value operator used in (2.3) preserves this relation, so that one obtains

(2.6) $$E[\min(n_1, n_2)] + E[\max(n_1, n_2)] = N.$$

Thus (2 6) and (2 2) could have been used in evaluating some of the forms above, or can be used as a check on the evaluations.

To compute the variance

$$\sigma^2(x) = \frac{\Sigma[x - E(x)]^2 f(x)}{N} = E[x^2] - \{E[x]\}^2, \tag{2.7}$$

it will be convenient to note that for the binomial case

$$\sigma^2_{\max} = \sigma^2_{\min} \tag{2.8}$$

where

$$\sigma^2_{\binom{\max}{\min}} = E\left[\begin{matrix}\max\\ \min\end{matrix}(n_1^2, n_2^2)\right] - \left\{E\left[\begin{matrix}\max\\ \min\end{matrix}(n_1, n_2)\right]\right\}^2, \tag{2.9}$$

and where because of the non-negative character of n_1 and n_2

$$E\left[\left\{\begin{matrix}\max\\ \min\end{matrix}(n_1, n_2)\right\}^2\right] = E\left[\begin{matrix}\max\\ \min\end{matrix}(n_1^2, n_2^2)\right].$$

To prove (2.8), note that for this binomial case

$$\{\max(n_1, n_2) - E[\max(n_1, n_2)]\}^2 = \{\min(n_1, n_2) - E[\min(n_1, n_2)]\}^2,$$

and thus each term for $\sigma^2_{\max}$ has its counterpart for $\sigma^2_{\min}$ when using the first part of (2.7) to compute these variances, and hence (2.8) must be true.

Defining σ^2 as the common value, one gets

$$\begin{aligned} 2\sigma^2 &= \sigma^2_{\max} + \sigma^2_{\min} \\ &= E[\max(n_1^2, n_2^2)] + E[\min(n_1^2, n_2^2)] - \{E[\max(n_1, n_2)]\}^2 \\ &\qquad - \{E[\min(n_1, n_2)]\}^2. \end{aligned} \tag{2.10}$$

The value of the sum

$$E[\max(n_1^2, n_2^2)] + E[\min(n_1^2, n_2^2)]$$

is somewhat easier to obtain than that of either part. For, max (n_1^2, n_2^2) is one of the integers (n_1^2, n_2^2) and min (n_1^2, n_2^2) is the other integer. Linearity in the expected value form then gives

$$\begin{aligned} E[\max(n_1^2, n_2^2)] + E[\min(n_1^2, n_2^2)] &= E[n^2 + (N - n)^2] \\ &= N^2p^2 + 2Np(1 - p) + N^2(1 - p)^2, \end{aligned} \tag{2.11}$$

a relation which is similar to (2.6).

Likewise one gets

$$\begin{aligned} &\{E[\max(n_1, n_2)]\}^2 + \{E[\min(n_1, n_2)]\}^2 \\ &\quad = \{E[\max(n_1, n_2)] + E[\min(n_1, n_2)]\}^2 \\ &\qquad\qquad - 2E[\max(n_1, n_2)]E[\min(n_1, n_2)] \\ &\quad = N^2 - 2E[\max(n_1, n_2)]E[\min(n_1, n_2)]. \end{aligned} \tag{2.12}$$

Substituting the results of (2.11) and (2.12) into (2.10), and solving for σ^2 one gets

$$
\begin{aligned}
\sigma^2 &= E[\max(n_1, n_2)]E[\min(n_1, n_2)] - N(N-1)p(1-p) \\
&= E[\max(n_1, n_2)]\{N - E[\max(n_1, n_2)]\} - N(N-1)p(1-p) \\
&= E[\min(n_1, n_2)]\{N - E[\min(n_1, n_2)]\} - N(N-1)p(1-p).
\end{aligned} \tag{2.13}
$$

If one desires, one can make independent evaluations of $E[\max(n_1^2, n_2^2)]$ and $E[\min(n_1^2, n_2^2)]$ and compute the variances from relation (2 9). Such evaluations bring into play the incomplete Beta functions at four different sets of values, with separate sets for N odd and N even Relations (2 13) seem preferable to this suggested "strong-arm" procedure. A proof of relation (2.8) by this means seems to be unduly algebraically complicated.

3. Normal approximation to the binomial distribution. If numerical values for large N are desired (beyond the range of tabulated values of the incomplete Beta Function) an approximation based on the normal distribution may be used. Let

$$
\begin{aligned}
n_1 &= Np_1 + x, \\
n_2 &= N - n_1 = N(1-p_1) - x,
\end{aligned} \tag{3.1}
$$

where the subscripts may be dropped when not needed for clarity
Then one has

$$
E\begin{bmatrix}\max \\ \min\end{bmatrix}(n_1, n_2) \cong \int_{-\infty}^{\infty} \frac{\begin{bmatrix}\max \\ \min\end{bmatrix}(x + Np, N(1-p) - x)}{\sqrt{2\pi Np(1-p)}} \cdot \exp\left(\frac{-x^2}{2Np(1-p)}\right) dx. \tag{3.2}
$$

To evaluate the minimum approximation, note that there are two ranges

$$
-\infty < x < \frac{N}{2} - Np, \tag{a}
$$

in which range $\min(x + Np, N(1-p) - x) = x + Np$,

$$
\frac{N}{2} - Np < x < \infty, \tag{b}
$$

in which range $\min(x + Np, N(1-p) - x) = N(1-p) - x$. Defining

$$
A(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx, \tag{3.3}
$$

a tabulated function, the integrations may be evaluated as

$$
\begin{aligned}
(3.4)\qquad E[\min (n_1, n_2)] &\cong NpA(M) + N(1-p)[1 - A(M)] \\
&\quad - \sqrt{\frac{2Np(1-p)}{\pi}} \exp\left[\frac{-N(1-2p)^2}{8p(1-p)}\right], \\
E[\max (n_1, n_2)] &\cong N(1-p)A(M) + Np[1 - A(M)] \\
&\quad + \sqrt{\frac{2Np(1-p)}{\pi}} \exp\left[\frac{-N(1-2p)^2}{8p(1-p)}\right],
\end{aligned}
$$

where

$$M = \frac{N/2 - Np}{\sqrt{Np(1-p)}}.$$

Note also that (2.6) holds for these approximate evaluations

For the variance, approximations (3.4) may be used in relations (2.13). Or, alternately, the variances may be computed by "strong-arm" methods using the definition (2.9). In this case, using the averaging defined implicitly by (2.10) one gets the evaluation

$$
\begin{aligned}
\sigma^2 &\cong N^2A(M)[1 - A(M)][1 - 2p]^2 + Np(1-p) \\
(3.5)\qquad &\quad + N(1-2p)[1 - 2A(M)] \sqrt{\frac{2Np(1-p)}{\pi}} \exp\left[\frac{-N(1-2p)^2}{8p(1-p)}\right] \\
&\quad - \frac{2Np(1-p)}{\pi} \exp\left[\frac{-N(1-2p)^2}{4p(1-p)}\right].
\end{aligned}
$$

It would seem preferable to use relations (2 13) rather than the above, for that reason the evaluation of forms (2.9) have not been included here.

4. Trinomial distributions. The form

$$(4\ 1)\qquad E\left[\begin{matrix}\max\\ \min\end{matrix}(n_1, n_2)\right] = \sum_{n_1+n_2+n_3=N} \frac{N!}{n_1!\,n_2!\,n_3!}\left[\begin{matrix}\max\\ \min\end{matrix}(n_1, n_2)\right] p_1^{n_1} p_2^{n_2} p_3^{n_3}$$

may be approximated, for large N, by the bivariate normal distribution. Suppose two attributes P (and not $P = \bar{P}$) and R (and not $R = \bar{R}$) are being observed in a distribution Then the four possible outcomes of an experiment could be represented as the categories PR, $P\bar{R}$, $\bar{P}R$, $\bar{P}\bar{R}$ with respective probabilities a, b, c, d; $a + b + c + d = 1$. In such a situation, for large N, one may use a bivariate normal distribution as a limiting form of the above described bivariate binomial distribution, or multinomial distribution with four categories

If the probability of one category, say PR, is zero, the bivariate normal distribution can be regarded as a limiting form of a trinomial distribution.

Indeed, defining

$$(4\ 2)\qquad x_1 = \frac{n_1 - Np_1}{[Np_1(1-p_1)]^{\frac{1}{2}}}; \qquad x_2 = \frac{n_2 - Np_2}{[Np_2(1-p_2)]^{\frac{1}{2}}},$$

the bivariate normal distribution takes the form [1]

$$(4.3)\quad dF = \frac{1}{2\pi(1-r^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1-r^2)}(x_1^2 - 2rx_1x_2 + x_2^2)\right\} dx_1\, dx_2,$$

where

$$-\infty < x_1, x_2 < \infty,$$

$$r = -\left[\frac{p_1 p_2}{(1-p_1)(1-p_2)}\right]^{\frac{1}{2}}.$$

The expected values are then given approximately by

$$(4.4)\qquad E\left[\begin{matrix}\max\\ \min\end{matrix}(n_1, n_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\begin{matrix}\max\\ \min\end{matrix}(n_1, n_2)\right] dF.$$

For the special case $p_1 = p_2$, evaluations have been made of $E[{}^{\max}_{\min}(n_1, n_2)]$ by the authors. For the finite summation (4.1), powers of N less than the one-half power were neglected, and the values

$$(4.5)\qquad \begin{aligned} E[\min(n_1, n_2)] &= Np - \left(\frac{Np}{\pi}\right)^{\frac{1}{2}}, \\ E[\max(n_1, n_2)] &= Np + \left(\frac{Np}{\pi}\right)^{\frac{1}{2}} \end{aligned}$$

were obtained.

For the integral case, again for $p_1 = p_2 = p$ and hence for $r = -p/(1-p)$, the evaluation proceeds as follows In virtue of (4.2) and (4 3)

$$(4\ 6)\qquad \begin{aligned} E[\min(n_1, n_2)] &= Np + [Np(1-p)]^{\frac{1}{2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\min(x_1, x_2)]\, dF \\ &= Np + [Np(1-p)]^{\frac{1}{2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\min(x_1 - x_2, 0)]\, dF. \end{aligned}$$

It is convenient to introduce a rotation of axes in order to evaluate integral (4 6). Indeed, rotation through $\pi/4$ radians will give

$$(4.7)\qquad \begin{aligned} x_1 &= \frac{y_1}{\sqrt{2}} - \frac{y_2}{\sqrt{2}}, \\ x_2 &= \frac{y_1}{\sqrt{2}} + \frac{y_2}{\sqrt{2}}, \end{aligned}$$

with

$$(4.8)\qquad x_1^2 + \frac{2p}{1-p} x_1x_2 + x_2^2 = y_1^2\left(\frac{1}{1-p}\right) + y_2^2\left(\frac{1-2p}{1-p}\right),$$

$$(4.9)\qquad \min(x_1 - x_2, 0) = \min(-y_2\sqrt{2}, 0),$$

$$(4.10)\qquad J \equiv \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = 1$$

Thus integral (4.6) becomes

$$
\begin{aligned}
&E[\min (n_1, n_2)] \\
&= Np + \left[\frac{Np(1-p)^3}{1-2p}\right]^{\frac{1}{2}} \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \\
&\cdot \exp\left[-\frac{1}{2}\frac{(1-p)^2}{1-2p}\left(\frac{y_1^2}{1-p} + y_2^2 \frac{(1-2p)}{1-p}\right)\right] \min(-y_2\sqrt{2}, 0)\, dy_1\, dy_2 \qquad (4.11)\\
&= Np + \left[\frac{Np(1-p)^3}{2(1-2p)}\right]^{\frac{1}{2}} \frac{1}{\pi} \int_{-\infty}^{\infty} \\
&\qquad \cdot \left\{\int_0^{\infty} - y_2 \exp\left[-\frac{1}{2}\frac{(1-p)}{1-2p} y_1^2 - \frac{(1-p)}{2} y_2^2\right] dy_2\right\} dy_1 .
\end{aligned}
$$

As indicated above, it is convenient to consider the form as an iterated integral, and integrate first with respect to y_2. The evaluation of (4.11) presents no serious difficulties,

$$
\begin{aligned}
E[\min (n_1, n_2)] &= Np - \left[\frac{Np(1-p)^3}{2(1-2p)}\right]^{\frac{1}{2}} \frac{1}{\pi(1-p)} \int_{-\infty}^{\infty} \\
&\qquad \cdot \exp\left[-\frac{(1-p)}{2(1-2p)} y_1^2\right] dy_1 \qquad (4.12)\\
&= Np - \left(\frac{Np}{\pi}\right)^{\frac{1}{2}}.
\end{aligned}
$$

Likewise

$$E[\max (n_1, n_2)] = Np + \left(\frac{Np}{\pi}\right)^{\frac{1}{2}}.$$

Note that these values are the same as those obtained from the finite summation form (4.1), as given by (4.5).

To evaluate the variance

$$\sigma^2 = E\left[\begin{matrix}\max \\ \min\end{matrix} (n_1^2, n_2^2)\right] - \left\{Np \pm \left(\frac{Np}{\pi}\right)^{\frac{1}{2}}\right\}^2 \qquad (4.13)$$

a finite summation form similar to (4.1) or an integral form similar to (4.4) may be used.

In case the integral form is used, it is convenient to introduce the variables x_1 and x_2 as defined by (4.2). One then gets

$$
\begin{aligned}
E[\min (n_1^2, n_2^2)] &= N^2p^2 + Np(1-p) \\
&\quad \cdot E\left[\min\left(x_1^2 + 2\left[\frac{Np}{1-p}\right]^{\frac{1}{2}} x_1 ; x_2^2 + 2\left[\frac{Np}{1-p}\right]^{\frac{1}{2}} x_2\right)\right] \qquad (4.14)\\
&= N^2p^2 + Np(1-p) + Np(1-p) \\
&\quad \cdot E\left[\min\left(x_1^2 - x_2^2 + 2\left[\frac{Np}{1-p}\right]^{\frac{1}{2}} (x_1 - x_2); 0\right)\right],
\end{aligned}
$$

in which one integration over the whole space has been carried out. Rotating axes as per (4.7) one gets

$$E[\min\,(n_1^2,\, n_2^2)] = N^2p^2 + Np(1-p) + 2Np(1-p) \cdot E\left[\min\left(-y_1y_2 - \left[\frac{2Np}{1-p}\right]^{\frac{1}{2}} y_2\,;\,0\right)\right]. \tag{4.15}$$

In evaluating this last expected value form, the region of integration may be considered as a sum of separate regions. Over some regions the integrand is zero, in other regions the non-negative product

$$y_2\left\{y_1 + \left[\frac{2Np}{1-p}\right]^{\frac{1}{2}}\right\}$$

is the integrand and this condition gives

$$\begin{cases} y_2 \geqq 0, \\ y_1 \geqq -\left[\dfrac{2Np}{1-p}\right]^{\frac{1}{2}}, \end{cases} \quad \text{and} \quad \begin{cases} y_2 \leqq 0, \\ y_1 \leqq -\left[\dfrac{2Np}{1-p}\right]^{\frac{1}{2}}, \end{cases}$$

as the regions of integration with the non-negative product as integrand.

Since the assumption that N is large has already been made, it is convenient to approximate further here and assume $[2Np/(1-p)]^{\frac{1}{2}}$ is large, and in particular to assume that integration from $-[2Np/(1-p)]^{\frac{1}{2}}$ to $+\infty$ is equal to integration from $-\infty$ to $+\infty$ for the integrand under consideration and for iterated integration with respect to the variable y_1.

Remark: An equivalent assumption is needed in the finite summation case when approximating $(Np)!$ by the use of Stirling's formula.

Thus one gets (since one of the above regions of integration is to be neglected)

$$\begin{aligned} E&\left[\min\left\{-y_2\left(y_1 + \left[\frac{2Np}{1-p}\right]^{\frac{1}{2}}\right);0\right\}\right] \\ &\cong \frac{-(1-p)}{2\pi(1-2p)^{\frac{1}{2}}} \int_{-\infty}^{\infty}\left[\int_0^{\infty} y_2\left(y_1 + \left[\frac{2Np}{1-p}\right]^{\frac{1}{2}}\right)\right. \\ &\qquad \left. \cdot \exp\left\{\frac{-(1-p)}{2(1-2p)}\, y_1^2 - \frac{(1-p)}{2}\, y_2^2\right\} dy_2\right] dy_1 \\ &= \frac{-1}{2\pi(1-2p)^{\frac{1}{2}}} \int_{-\infty}^{\infty}\left(y_1 + \left[\frac{2Np}{1-p}\right]^{\frac{1}{2}}\right) \exp\left[\frac{-(1-p)}{2(1-p)}\, y_1^2\right] dy_1 \\ &= -\frac{1}{1-p}\left(\frac{Np}{\pi}\right)^{\frac{1}{2}}. \end{aligned} \tag{4.16}$$

Collecting results from (4.13), (4.15) and (4.16) one obtains

$$\sigma_{\min}^2 \cong Np\left(1 - p - \frac{1}{\pi}\right). \tag{4.17}$$

By a similar procedure, one may compute also that

$$\sigma_{\max}^2 \cong Np\left(1 - p - \frac{1}{\pi}\right). \tag{4.18}$$

For this three category case, the proof used to obtain relation (2 8) is no longer applicable, yet the relation $\sigma^2_{\min} = \sigma^2_{\max}$ still holds for the approximating relations given above

5. Conclusion. Since the normal distribution was used in some instances to obtain approximations for the binomial and multinomial distributions, many of the maximum and minimum relations stated as approximations for the multinomial are exact for the appropriate normal distribution.

No convenient formulation was found for the general trinomial case (p_1, p_2, p_3 unequal) similar to relations (4.5), (4.17), and (4.18).

As possible applications of the general solution of this problem, the referee has kindly supplied the authors with a reference of Guttman [2]. Sampling theory provided by the general solution to this problem could be used in connection with Guttman's reliability coefficient.

REFERENCES

[1] M. G KENDALL, *The Advanced Theory of Statistics*, Vol I, 3rd edition, Charles Griffin and Co., 1947, p 133.

[2] LOUIS GUTTMAN, "The test-retest reliability of qualitative data," *Psychometrika*, Vol. 11 (1946), pp 81–95

DERIVATION OF A BROAD CLASS OF CONSISTENT ESTIMATES

By R. C. Davis

U. S. Naval Ordnance Test Station, Inyokern, California

1. Summary. Given a chance vector $\mathbf{X}$ with distribution function $F(\mathbf{X}, \boldsymbol{\theta}_T)$, where $\boldsymbol{\theta}_T$ denotes the true unknown parameter vector, a broad class of estimates of $\boldsymbol{\theta}_T$ is derived which is shown to be identical with the class of all consistent estimates of $\boldsymbol{\theta}_T$. A sub-class is obtained each member of which has the following properties: a) Its construction depends upon the solution of an equation involving a single vector function of the parameter vector $\boldsymbol{\theta}$ and the members of a sequence $\{\mathbf{X}_n\}$ of independent and identically distributed chance vectors; b) the estimate so obtained converges almost certainly to $\boldsymbol{\theta}_T$, c.) it is a symmetric function of the members of the sequence $\{\mathbf{X}_n\}$. In order to obtain this sub-class it is postulated that a function of $\mathbf{X}$ and $\boldsymbol{\theta}$ exists (continuous in $\boldsymbol{\theta}$ for a certain neighborhood of the true parameter $\boldsymbol{\theta}_T$ and existing for each $\mathbf{X}$ in a subset of the sample space) which satisfies a Lipschitz condition in $\boldsymbol{\theta}$. In particular if a density function $f(\mathbf{X}, \boldsymbol{\theta}_T)$ exists satisfying certain conditions, the consistency of the maximum likelihood estimate can be established under regularity conditions quite different from those usually assumed [1] This is not to be interpreted as a weakening of the usual regularity conditions but rather as an extension of the class of consistent likelihood estimates obtained under the usual regularity conditions

2. Introduction. The present work is the result of investigations into the following question posed by J. Neyman: What happens to the asymptotic properties of the maximum likelihood estimate of $\boldsymbol{\theta}_T$ when the usual regularity conditions on $F(\mathbf{X}, \boldsymbol{\theta})$ are relaxed? The consistency and efficiency of the estimate are the properties in question, and the present work arose from the observation that consistency at least can be obtained under conditions much different than those usually assumed [1]. The assumptions made below are existential in nature, and no general methods are given for the actual construction of consistent estimates. As stated above, however, the results of this work can be used to widen the class of consistent maximum likelihood estimates established heretofore. Although simple upper and lower bounds for the variance of a consistent estimate are obtained, no answer is given to the question of determining the efficiency of such an estimate. In regard to consistent estimates, J. Neyman and E Scott have discussed recently [2] the need for a systematic method of obtaining consistent estimates. Wald has given necessary and sufficient conditions [3] for the existence of a uniformly consistent estimate of an unknown parameter $\boldsymbol{\theta}$ when there exists a density function continuous jointly in all of its arguments, and it is assumed that the domain of each of the unknown parameters is a closed and bounded set. It is hoped that the class of consistent estimates

derived below will help shed some light on a general method for actually obtaining such estimates. In this connection it is important to point out that if necessary and sufficient conditions were known for the existence and uniqueness of a fixed point for a transformation on E_n to E_n, the weakest possible conditions could be expressed for the existence of consistent estimates obtained in the manner given below. It is surmised that the use of a Holder condition of order one as presented below is stronger than required.

Let $\{\mathbf{X}_i\}$, $i = 1, 2, \cdots, n, \cdots$, be a sequence of chance vectors in which $\mathbf{X}_i$ possesses the probability distribution function $F_i(\mathbf{X}, \boldsymbol{\theta})$ depending upon an unknown parameter vector $\boldsymbol{\theta}$. The vector $\mathbf{X}$ has components X_i, $i = 1, 2, \cdots, s$, where X_i is a chance variable, and $\boldsymbol{\theta}$ has components θ_j, $j = 1, 2, \cdots, m$. The problem is to obtain a function of the $\mathbf{X}_i$ which is a consistent estimate of $\boldsymbol{\theta}$ We denote by E_s the real Euclidean space of s dimensions and by E_s' a subset of E_s excluding at most a set of probability measure zero. For convenience we use the symbol $\|\theta\|$ to denote the norm of $\boldsymbol{\theta}$, where

$$\|\boldsymbol{\theta}\| = (\theta_1^2 + \theta_2^2 + \cdots + \theta_m^2)^{1/2}.$$

We define in a similar manner the norm of any function which assumes values in E_m. The following assumption is made:

ASSUMPTION 1. *There exists a point $\boldsymbol{\theta}_0$ and a neighborhood $W(\boldsymbol{\theta}_0, a)$ of $\boldsymbol{\theta}_0$ having radius a $(a > 0)$ which contains the true parameter vector $\boldsymbol{\theta}_T$ as an interior point and there exists an infinite sequence of functions $G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta})$, $n = 1, 2, \cdots$, ad inf. on $E_s \times E_m$ to E_m such that*

(a) *for each n the equation*

$$G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta}) = 0$$

has a unique solution $\boldsymbol{\theta} = \boldsymbol{\theta}_n^*(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ in $W(\boldsymbol{\theta}_0, a)$. (For the sake of brevity we usually write $G_n(\mathbf{X}; \boldsymbol{\theta}) = G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta})$.)

(b) *For every pair of values of $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ in $W(\boldsymbol{\theta}_0, a)$ and for some K with $0 < K < 1$*

$$\lim_{n\to\infty} P\{\|G_n(\mathbf{X}, \boldsymbol{\theta}_1) - G_n(\mathbf{X}, \boldsymbol{\theta}_2) - (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)\| \leqq K \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|\} = 1.$$

(c) *For every $\epsilon > 0$,*

$$\lim_{n\to\infty} P\{\|G_n(\mathbf{X}, \boldsymbol{\theta}_T)\| < \epsilon\} = 1.$$

3. A consistent estimate of $\boldsymbol{\theta}_T$.

THEOREM 3.1. *The solution $\boldsymbol{\theta} = \boldsymbol{\theta}_n^*(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ of the equation*

$$G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \boldsymbol{\theta}) = 0$$

is a consistent estimate of $\boldsymbol{\theta}_T$, providing $G_n(\mathbf{X}; \boldsymbol{\theta})$ satisfies Assumption 1.

PROOF: From Assumption 1b it follows that given $\delta > 0$, we have for all $n > N'(\delta)$,

$$P\{\|G_n(\mathbf{X}, \boldsymbol{\theta}_T) - (\boldsymbol{\theta}_T - \boldsymbol{\theta}_n^*)\| \leqq K \|\boldsymbol{\theta}_T - \boldsymbol{\theta}_n^*\|\} > 1 - \frac{\delta}{2}, \tag{3.1}$$

since $G_n(\mathbf{X}, \theta_n^*) = 0$. It follows from (3.1) that for all $n > N'(\delta)$,

$$(3.2) \qquad P\left\{\frac{\| G_n(\mathbf{X}, \theta_T) \|}{1+K} \leqq \| \theta_T - \theta_n^* \| \leqq \frac{\| G_n(\mathbf{X}, \theta_T) \|}{1-K}\right\} > 1 - \frac{\delta}{2}.$$

From Assumption 1c it follows that there exists $N''(\epsilon, \delta)$ such that $n > N''(\epsilon, \delta)$ implies

$$(3.3) \qquad P\{\| G_n(\mathbf{X}, \theta_T) \| < \epsilon(1 - K)\} > 1 - \frac{\delta}{2}.$$

(3.2), (3.3), and a familiar formula in probability imply for all

$$n > \max\,[N'(\delta), N''(\epsilon, \delta)],$$
$$P\{\| \theta_T - \theta_n^* \| < \epsilon\} > 1 - \delta.$$

It is noted that (3.2) characterizes the speed of convergence of the estimate θ_n^*. The following uniqueness property is noted: *If a given sequence of functions $G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \theta)$ satisfies Assumption* 1, *then θ_T is the unique parameter vector in $W(\theta_0, a)$ which satisfies item* c *of Assumption* 1. The proof of this remark is left to the reader.

The following remark demonstrates the extreme generality of the class of consistent estimates obtained in the above manner: *The set of estimates of the parameter vector θ_T obtained from the class of all sequences of functions*

$$G_n(\mathbf{X}_1, \mathbf{X}_2, \cdot \cdot, \mathbf{X}_n; \theta)$$

satisfying Assumption 1 *is identical with the set of all consistent estimates of the parameter vector θ_T*. The proof of this remark is quite obvious and is left to the reader.

4. Properties of a sub-class of consistent estimates. The question arises naturally concerning a general method for the construction of a sequence of functions $G_n(\mathbf{X}_1, \mathbf{X}_2, \cdot \cdot, \mathbf{X}_n, \theta)$ satisfying Assumption 1. The author knows of no general method. It is possible to describe a sub-class of the class of consistent estimates, the construction of which depends upon the existence of one function rather than a sequence of functions. This is possible by application of the strong law of large numbers, and in this way consistent estimates of the parameter vector are obtained which converge almost certainly to the true value θ_T Moreover it is clear that under certain conditions the function

$$G_n(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n; \theta_T)$$

defined as in equation 4.1 below is an asymptotically m-variate normal variable

ASSUMPTION 2. *Let $\{\mathbf{X}_i\}$, $i = 1, 2, \cdots, n, \cdot\cdot$, be a sequence of independently and identically distributed chance vectors with common distribution function $F(\mathbf{X}; \theta)$, where θ is again the unknown parameter vector.*

ASSUMPTION 3. *There exists a function $g(\mathbf{X}, \theta)$ on $E_s \times E_m$ to E_m such that* (a) *for every $\mathbf{X} \in E_s'$ and every distinct pair (θ_1, θ_2) in $W(\theta_0, a)$,*

$$\| g(\mathbf{X}, \theta_1) - g(\mathbf{X}, \theta_2) - (\theta_1 - \theta_2) \| \leqq K \| \theta_1 - \theta_2 \|,$$

where $0 < K < 1$ *and* $\| g(\mathbf{X}, \theta_0) \| < (1 - K)a$.

(b) $Eg(\mathbf{X}, \theta_T) = \int_{-\infty}^{\infty} g(\mathbf{X}, \theta_T)\, dF(\mathbf{X}, \theta_T) = 0.$

We define the function $G_n(\mathbf{X}, \theta)$ as follows:

$$G_n(\mathbf{X}, \theta) = \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{X}_i, \theta). \tag{4.1}$$

The following lemmas are required:

LEMMA 4.1. *$G_n(\mathbf{X}, \theta)$ as defined in* (4.1) *satisfies the conditions in Assumption* 3 *with $G_n(\mathbf{X}, \theta)$ replacing $g(\mathbf{X}, \theta)$.*

The proof is sufficiently obvious to be omitted

LEMMA 4.2 *$G_n(\mathbf{X}, \theta_T) \to 0$ almost certainly as $n \to \infty$, if Assumptions* 2 *and* 3b *hold*

PROOF: Since $Eg(\mathbf{X}_i, \theta_T) = 0$, $i = 1, 2, \cdots, n$, and the chance variables $g(\mathbf{X}_i, \theta_T)$ are independently and identically distributed, this follows immediately from a theorem due to Kolmogorov [5].

THEOREM 4.1. *If Assumptions* 2 *and* 3 *hold, then the equation $G_n(\mathbf{X}, \theta) = 0$ has a unique solution $\theta = \theta_n^*(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ in $W(\theta_0, a)$, where θ_n^* is a consistent estimate of θ_T and is moreover a symmetric function of the observation vectors $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$*

PROOF. We obtain the solution θ_n^* by the method of successive substitutions Define

$$\theta_1 = \theta_0 - G_n(\mathbf{X}, \theta_0), \quad \cdots, \theta_{q+1} = \theta_q - G_n(\mathbf{X}, \theta_q).$$

In view of Lemma 4.1 we can apply a well known existence theorem [4] in the theory of functions to prove that the sequence $\{\theta_q\}$ converges to a limit θ_n^* which is also in $W(\theta_0, a)$. The same theorem establishes the uniqueness of the solution in $W(\theta_0, a)$ This uniqueness property together with lemmas 4.1 and 4.2 establish the fact that the sequence $\{G_n(\mathbf{X}, \theta)\}$ as defined in equation (4.1) satisfies Assumption 1 It follows immediately from Theorem 3.1 that θ_n^* is a consistent estimate of θ_T We can, however, prove a stronger relationship

THEOREM 4.2 *The estimate θ_n^* defined in Theorem* 4.1 *converges almost certainly to θ_T.*

PROOF: From Lemma 4.2 we know that given any number $\epsilon > 0$, there exists an integer $N(\epsilon)$ such that for all $n > N(\epsilon)$

$$P\{\| G_n(\mathbf{X}, \theta_T) \| < \epsilon(1 - K)\} = 1.$$

From Assumption 3a and Lemma 4.1 we see that

$$\| G_n(\mathbf{X}, \theta_T) - (\theta_T - \theta_n^*) \| \leqq K \| \theta_T - \theta_n^* \|,$$

since $G_n(\mathbf{X}, \theta_n^*) = 0$ Then

$$\| G_n(\mathbf{X}, \theta_T) \| \geqq (1 - K) \| \theta_T - \theta_n^* \|.$$

Clearly the set of $\mathbf{X} \in E_s'$ for which $\| \theta_T - \theta_n^* \| < \epsilon$ includes the set of $\mathbf{X}$ for which $\| G_n(\mathbf{X}, \theta_T) \| < \epsilon(1 - K)$.

Therefore, for $n > N(\epsilon)$,

$$P\{\| \theta_T - \theta_n^* \| < \epsilon\} \geqq P\{\| G_n(\mathbf{X}, \theta_T) \| < \epsilon(1 - K)\} = 1,$$

and the proof is completed.

The uniqueness of the parameter value θ_T in the neighborhood $W(\theta_0, a)$ follows immediately from the remark succeeding Theorem 3.1 since Assumption 1 is valid in Theorems 4.1 and 4.2.

It is interesting to note that the application of a theorem in the theory of functions of a real variable gives the result that if the function $g(\mathbf{X}, \theta)$ is continuous on a bounded and closed set in $E_s \times E_m$ and if we take for E_s' a bounded and closed set, then $\theta_n^*(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ is a continuous function of $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ for $\mathbf{X}_i \in E_s'$ $(i = 1, 2, \cdots, n)$. If we assume the continuity of $g(\mathbf{X}, \theta)$ in $\mathbf{X}$ for each θ in $W(\theta_0, a)$ the following remark demonstrates an interesting relationship concerning the uniqueness of the solution for θ in the equation $Eg(\mathbf{X}, \theta) = 0$. *If in addition to Assumption 3 we assume that $g(\mathbf{X}, \theta)$ is continuous in $\mathbf{X}$ for every $\mathbf{X}$ in E_s and every θ in $W(\theta_0, a)$ and if at least one of the components $g_i(\mathbf{X}, \theta)$, $1 \leqq i \leqq m$ of the m-dimensional vector function $g(\mathbf{X}, \theta)$ satisfies also a Lipschitz condition:*

$$\| g_i(\mathbf{X}, \theta_1) - g_i(\mathbf{X}, \theta_2) - (\theta_1 - \theta_2) \| \leqq K \| \theta_1 - \theta_2 \|$$

for every distinct pair θ_1, θ_2 in $W(\theta_0, a)$, then for all θ in $W(\theta_0, a)$, θ_T is the unique solution for θ of the equation $Eg(\mathbf{X}, \theta) = 0$.

The proof of this remark is left to the reader.

5. Upper and lower bounds for the expected squared error of $\theta_n^*(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$. Denote by $g_i(\mathbf{X}, \theta)$, $i = 1, 2, \cdots, m$, the m components of the chance vector $g(\mathbf{X}, \theta)$. We now make an additional assumption.

ASSUMPTION 4.

$$E[g_i(\mathbf{X}, \theta_T)g_j(\mathbf{X}, \theta_T)] = \lambda_{ij}$$

exists for $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, m$

It follows from Assumptions 2, 3b, and 4 and the Lindeberg-Lévy form of the Central Limit Theorem that the vector $\sqrt{n}G_n(\mathbf{X}, \theta_T)$ tends in probability to an m-variate normal distribution with means zero and moment matrix (λ_{ij}).

Now from Assumption 3a and Lemma 4.1

$$\frac{\| G_n(\mathbf{X}, \theta_T) \|}{1 + K} \leqq \| \theta_n^* - \theta_T \| \leqq \frac{\| G_n(\mathbf{X}, \theta_T) \|}{1 - K} \tag{5.1}$$

For convenience define

$$\lambda = \sum_{i=1}^{m} \lambda_{ii}.$$

We obtain then

$$E\,\|\,G_n(\mathbf{X},\,\theta_T)\,\|^2 = \frac{\lambda}{n}.$$

It follows then from equation (5.1) that

$$\frac{\lambda}{n(1+K)^2} \leq E\,\|\,\theta_n^* - \theta_T\,\|^2 \leq \frac{\lambda}{n(1-K)^2}.$$

6. The consistency of maximum likelihood estimates. The results of this paper can be used to extend the class of consistent maximum likelihood estimates established heretofore [1].[1] Assume that $F(\mathbf{X},\,\theta)$ admits a density function $f(\mathbf{X},\,\theta)$ with the property

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(\mathbf{X},\theta)\,d\mathbf{X} = \int_{-\infty}^{\infty}\frac{\partial f}{\partial\theta}\,(\mathbf{X},\theta)\,d\mathbf{X}.$$

Then

$$E\left[\frac{\partial}{\partial\theta}\ln f(\mathbf{X},\theta)\right] = 0.$$

The maximum likelihood estimate of θ_T is obtained by solving the equation

$$\frac{\partial}{\partial\theta}\ln L(\mathbf{X},\theta) = 0,$$

where

$$L(\mathbf{X},\theta) = \prod_{i=1}^{n} f(\mathbf{X}_i\,,\theta).$$

If a sample $\mathbf{X}_1\,,\ \mathbf{X}_2\,,\ \cdots,\ \mathbf{X}_n$ is obtained as the result of n random independent drawings from the distribution having the c.d.f. $F(\mathbf{X},\,\theta)$, the sample values will satisfy Assumption 2. Assumption 3b holds as assumed above If we assume also that the function $\partial/\partial\theta \ln f(\mathbf{X},\,\theta)$ satisfies Assumption 3a, it follows directly from Theorem 4.2 that the maximum likelihood estimate converges almost certainly to the true parameter vector as the sample size approaches infinity.

The author wishes to acknowledge his indebtedness and gratitude to Professor Jerzy Neyman for the many helpful suggestions made during the preparation of the paper.

REFERENCES

[1] J. L. Doob, "Probability and statistics," *Trans. Am. Math. Soc.*, Vol. 36 (1934), p. 759.

[2] J Neyman and Elizabeth L. Scott, "Consistent estimates based on partially consistent observations," *Econometrica*, Vol 16 (1948), pp. 1–32

[3] A. Wald, "Estimation of a parameter when the number of unknown parameters increases indefinitely with the number of observations," *Annals of Math. Stat.*, Vol 19 (1948), pp. 220–227.

[1] Recently Wald [6] and Wolfowitz [7] have discussed the consistency of the maximum likelihood estimate from another approach than the one employed by Doob.

[4] L. M Graves, *The Theory of Functions of Real Variables*, McGraw-Hill Book Co , 1946.
[5] A Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Chelsea Publishing Co , 1946
[6] A Wald, "Note on the consistency of the maximum likelihood estimate," *Annals of Math Stat* , Vol. 20 (1949), pp. 595-600
[7] J. Wolfowitz, "On Wald's proof of the consistency of the maximum likelihood estimate," *Annals of Math. Stat.*, Vol 20 (1949), pp. 601-602

DISTRIBUTION OF THE SUM OF ROOTS OF A DETERMINANTAL EQUATION UNDER A CERTAIN CONDITION

By D. N. Nanda

University of North Carolina

1. Summary. This paper is in continuation of the author's first two papers [1] and [2] In this paper a method is described by which it is possible to derive the distribution of the sum of roots of a certain determinantal equation under the condition that $m = 0$. This condition implies, when the results are applied to canonical correlations, that the numbers of variates in the two sets differ by unity. The distributions for the sum of roots under this condition have been obtained for $l = 2, 3$ and 4 and are given in this paper This paper also derives the moments of these distributions.

2. Introduction. The reader should refer to the first two papers of this series [1] and [2] for detailed explanation of the preliminaries essential for this paper.

The distribution of any root of the determinantal equation, specified by its rank when the roots are arranged in a descending order of magnitude, was derived by the author [1] The distribution of the largest root was expressed as

$$(1) \qquad P_r(\theta_1 \leq x) = C(l, m, n)F_{l,m,n}(x) = \text{const.}\ (0, l, l-1, \cdots, 1, x; m, n).$$

3. Method. Putting $\theta_i = \rho_i/n$ in $R(l, m, n)$ as given in [1] and allowing n to tend to infinity, the distribution density reduces to

$$R(l, m) = \text{const.}\ \Pi\rho_i^m \prod_{i<j} (\rho_i - \rho_j)e^{-\Sigma\rho_i} \qquad (0 < \rho_l < \rho_{l-1} < \cdots < \rho_1 < \infty),$$

where the constant is independent of n, by [2]. If we replace x by x/n in the right-hand side of (1) and allow n to tend to infinity, then the resulting function $G_{l,m}(x)$ is independent of n and it can be shown by comparing the two methods A and B in [2], that

$$(2) \qquad \int_{0<\rho_l<\rho_{l-1}<\cdots<\rho_1<x} R(l, m)\Pi\, d\rho_i = G_{l,m}(x).$$

This is a constant multiple of

$$(3) \qquad \phi(x, m) = \int_{0<\rho_l<\rho_{l-1}<\cdots<\rho_1<x} \Pi\rho_i^m \prod_{i<j} (\rho_i - \rho_j)e^{-\Sigma\rho_i}\, \Pi\, d\rho_i$$

$$= \text{const.}\ x^{l+lm+l(l-1)/2}\,\theta(x, m).$$

Putting $\rho_i = xy_i$, we have

$$(4) \qquad \int_{0<y_l<y_{l-1}<\cdots<y_1<1} \Pi y_i^m \prod_{i<j} (y_i - y_j)e^{-x\Sigma y_i}\, \Pi\, dy_i = \text{const.}\ \theta(x, m)$$

The left-hand side is proportional to the moment generating function for the sum of roots when $n = 0$.

Let $y_1 = 1 - \theta_l$, $y_2 = 1 - \theta_{l-1}$, $\cdots$, $y_l = 1 - \theta_1$; then (4) gives

$$(5) \quad \int_{0<\theta_l<\theta_{l-1}<\quad<\theta_1<1} \Pi(1 - \theta_i)^m \prod_{i<j} (\theta_i - \theta_j) e^{-lx+x\Sigma\theta_i} \Pi\, d\theta_i = \text{const.}\, \theta(x, m).$$

Let m be changed to n and both sides be multiplied by e^{lx}, then we get

$$(6) \quad \int_{0<\theta_l<\theta_{l-1}<\cdots<\theta_1<1} \Pi(1 - \theta_i)^n \prod_{i<j} (\theta_i - \theta_j) e^{x\Sigma\theta_i} \Pi d\theta_i = \text{const}\ e^{lx}\, \theta(x, n).$$

The left-hand side of (6) is the moment generating function for the sum of roots when $m = 0$.

The method for obtaining the probability distributions is described in detail for each of the cases $l = 2, 3$, in the following sections.

It may, however, be added here that the condition $m = 0$, implies that $|\, p - q\, | = 1$ in the case of canonical correlations. It also implies, in generalized analysis of variance, that if we have K samples and measurements are made on p characters then $K - 1$ and p should differ by unity. Thus the distribution is given for 5 samples and 3 characters when $l = 3$ $(p = 3)$.

4. Distribution of the sum of roots when m = 0.

(*a*) $l = 2$. The value of $G_{2,m}(x)$ has been given in [2] as

$$(7) \qquad G_{2,m}(x) = k(2, m) \left[2 \int_0^x u^{2m+1} e^{-2u}\, du - x^{m+1} e^{-x} \int_0^x u^m e^{-u}\, du \right],$$

where $K(2, m) = 2^{2m+1}/\Gamma(2m + 2)$. Then in the notation just given

$$\phi(x, m) = 2 \int_0^x u^{2m+1} e^{-2u}\, du - x^{m+1} e^{-x} \int_0^x u^m e^{-u}\, du.$$

Replacing u by xu, we get

$$\begin{aligned} \phi(x, m) &= 2x^{2m+2} \int_0^1 u^{2m+1} e^{-2xu}\, du - x^{2m+2} e^{-x} \int_0^1 u^m e^{-xu}\, du \\ (8) \qquad &= \frac{x^{2m+2}}{m + 1} \int_0^1 e^{-2xu}\, d(u^{2m+2}) - \frac{x^{2m+2}\, e^{-x}}{m + 1} \int_0^1 e^{-xu}\, d(u^{m+1}) \\ &= \frac{x^{2m+3}}{m + 1} \left[2 \int_0^1 u^{2m+2} e^{-2xu}\, du - e^{-x} \int_0^1 u^{m+1} e^{-xu}\, du \right]. \end{aligned}$$

Hence

$$\theta(x, m) = \text{const.} \left[2 \int_0^1 u^{2m+2} e^{-2xu}\, du - e^{-x} \int_0^1 u^{m+1} e^{-xu}\, du \right],$$

and according to (6),

$$\int_{0<\theta_2<\theta_1<1} \Pi(1-\theta_i)^n(\theta_1-\theta_2)e^{x\Sigma\theta_i}\,d\theta_1\,d\theta_2$$

$$= \text{const. } e^{2x}\left[2\int_0^1 u^{2n+2}e^{-2xu}\,du - e^{-x}\int_0^1 u^{n+1}e^{-xu}\,du\right] \tag{9}$$

$$= \text{const.}\left[2\int_0^1 (1-u)^{2n+2}e^{2xu}\,du - \int_0^1 (1-u)^{n+1}e^{xu}\,du\right],$$

by replacing u by $1-u$. Or,

$$E(e^{x\Sigma\theta_i}) = \text{const.}\left[2\int_0^1 (1-u)^{2n+2}e^{2xu}\,du - \int_0^1 (1-u)^{n+1}e^{xu}\,du\right].$$

The constant can be evaluated by putting $x = 0$.

Then let $P_r(\theta_1+\theta_2 \le Z) = \text{const. } [F_1(Z) + F_2(Z)]$, where $F_1(Z)$ and $F_2(Z)$ are cumulative distribution functions given by integrating the density $(1-u)^{2n+2}$ of $2u$ and $(1-u)^{n+1}$ of u, respectively. It is easily seen that

$$F_2(Z) = \int_0^Z (1-u)^{n+1}\,du = [1-(1-Z)^{n+2}]/(n+2) \qquad (Z \le 1).$$

Since $F_1(Z)$ is to be obtained from the density of $2u$, we may substitute $v = 2u$ and then integrate. Thus

$$F_1(Z) = 2\int_0^Z \left(1-\frac{v}{2}\right)^{2n+2} dv/2 = 2[1-(1-Z/2)^{2n+3}]/(2n+3) \quad (Z \le 2).$$

Hence the result for $l = 2$ is

$$P_r(\theta_1+\theta_2 \le Z) = 2(n+2)[1-(1-Z/2)^{2n+3}] - (2n+3)[1-(1-Z)^{n+2}] \qquad (0 \le Z \le 1),$$

$$= 2(n+2)[1-(1-Z/2)^{2n+3}] - (2n+3) \qquad (1 \le Z \le 2).$$

(b) $l = 3$. The value of $G_{3,m}(x)$ as given in [2] is changed as

$$G_{3,m}(x) = K(3,m)\left\{2\int_0^x u^{2m+3}e^{-2u}\,du\int_0^x u^m e^{-u}\,du - 2\int_0^x u^{m+1}e^{-u}\,du \cdot \int_0^x u^{2m+2}e^{-2u}\,du - \frac{x^{m+2}e^{-x}}{m+1}\left[2x^{2m+3}\int_0^1 u^{2m+2}e^{-2xu}\,du - x^{2m+3}e^{-x} \cdot \int_0^1 u^{m+1}e^{-xu}\,du\right]\right\},$$

using (8). $K(3, m)$ is a constant independent of n. Putting xu for u in only the first two terms of the right-hand side of the above equation, we get

$$G_{3,m}(x) = k(3, m)x^{3m+5}\left\{2\int_0^1 u^{2m+3}e^{-2xu}\,du\int_0^1 u^m e^{-xu}\,du\right.$$

$$(10) \qquad - 2\int_0^1 u^{2m+2}e^{-2xu}\,du\int_0^1 u^{m+1}e^{-xu}\,du - \frac{2e^{-x}}{m+1}\int_0^1 u^{2m+2}e^{-2xu}\,du$$

$$\left. + \frac{e^{-2x}}{m+1}\int_0^1 u^{m+1}e^{-xu}\,du\right\}.$$

By integrating by parts we get x^{3m+6} as a common factor on the right-hand side of the above equation. Then according to (5) and (6) we have

$$\int_{0<y_3<y_2<y_1<1} \Pi y_i^m \prod_{i<j} (y_i - y_j)e^{-x\Sigma y_i}\,\Pi\, dy_i = \text{const.}\left\{2(m+2)\right.$$

$$\cdot \int_0^1 u^{2m+3}e^{-2xu}\,du\int_0^1 u^{m+1}e^{-xu}\,du + 2(2m+3)e^{-x}\int_0^1 u^{2m+4}e^{-2xu}\,du$$

$$\left. - 4(m+2)e^{-x}\int_0^1 u^{2m+3}e^{-2xu}\,du + e^{-2x}\int_0^1 u^{m+2}e^{-xu}\,du\right\}.$$

Putting $y_1 = 1 - \theta_3$, $y_2 = 1 - \theta_2$, $y_3 = 1 - \theta_1$ and, changing m to n and multiplying with e^{3x} we get

$$\int_{0<\theta_3<\theta_2<\theta_1<1} \Pi(1-\theta_i)^n \prod_{i<j} (\theta_i - \theta_j)e^{x\Sigma\theta_i}\,\Pi\, d\theta_i$$

$$(11) \qquad = \text{const.}\left\{2(n+2)\int_0^1 u^{2n+3}e^{2x(1-u)}\,du\int_0^1 u^{n+1}e^{x(1-u)}\,du\right.$$

$$+ 2(2n+3)\int_0^1 u^{2n+4}e^{2x(1-u)}\,du - 4(n+2)\int_0^1 u^{2n+3}e^{2x(1-u)}\,du$$

$$\left. + \int_0^1 u^{n+2}e^{x(1-u)}\,du\right\}.$$

Thus we have

$$P_r(\theta_1 + \theta_2 + \theta_3 \le Z) = \text{const.}\,\{F_1(Z) + F_2(Z) + F_3(Z) + F_4(Z)\},$$

where $F_1(Z)$, $F_2(Z)$, $F_3(Z)$ and $F_4(Z)$ are the contributions to the cumulative distribution by the four terms of the right-hand side of the following equation

$$E(e^{x\Sigma\theta_i}) = \text{const.}\left\{2(n+2)\int_0^1 (1-u)^{2n+3}e^{2xu}\,du\int_0^1 (1-u)^{n+1}e^{xu}\,du\right.$$

$$+ 2(2n+3)\int_0^1 (1-u)^{2n+4}e^{2xu}\,du - 4(n+2)\int_0^1 (1-u)^{2n+3}e^{2xu}\,du$$

$$\left. + \int_0^1 (1-u)^{n+2}e^{xu}\,du\right\},$$

where const. $= [(n+2)(n+3)(2n+5)]$. Proceeding according to the method given in (a) we have

$$(12) \qquad F_4(Z) = [1 - (1-Z)^{n+3}]/(n+3) \qquad (0 \le Z \le 1),$$

(13) $$F_2(Z) = 2(2n+3)[1-(1-Z/2)^{2n+5}]/(2n+5) \qquad (0 \le Z \le 2),$$

(14) $$F_3(Z) = -4(n+2)[1-(1-Z/2)^{2n+4}]/(2n+4) \qquad (0 \le Z \le 2).$$

Let us now consider $F_1(Z)$, which is the contribution of the first term. Let y_1 and y_2 be distributed between 0 and 1 with densities $(1-y_1)^{2n+3}$ and $(1-y_2)^{n+1}$

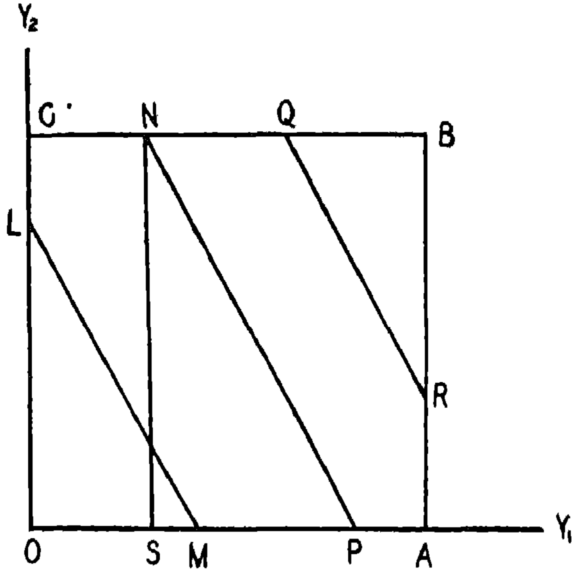


FIG. 1

respectively, then

$$F_1(Z) = 2(n+2) \iint_{2y_1+y_2 \le Z} (1-y_1)^{2n+3}(1-y_2)^{n+1}\, dy_1\, dy_2\,,$$

where Z goes from 0 to 3.

Let us consider the distribution over the unit square $OABC$, Fig. 1, then for $Z \le 1, Z \le 2$, and $Z \le 3$; we have to integrate over OLM, $OCNP$, and $OCQRA$, where LM, NP and QR are the three lines given by $2y_1 + y_2 \le Z$ according as $Z \le 1, Z \le 2$, and $Z \le 3$.

(i) The integration over OLM is given below

$$F_{1,1}(Z) = 2(2n+2) \iint_{2y_1+y_2 \le Z} (1-y_1)^{2n+3}(1-y_2)^{n+1}\, dy_1\, dy_2 \qquad \text{for } Z \le 1,$$

or

$$F_{1,1}(Z) = 2\left\{\frac{1}{2n+4}[1-(1-Z/2)^{2n+4}]\right.$$

(15) $$-\lambda \cdot 2^{n+2}\left(\frac{3-Z}{2}\right)^{3n+6} [I_{2/(3-Z)}(2n+4, n+3)$$

$$\left.-\, I_{(2-Z)/(3-Z)}(2n+4, n+3)]\right\},$$

where

$$\lambda = B(2n + 4, n + 3) = \int_0^1 y^{2n+3}(1 - y)^{n+2}\, dy$$

and

$$\lambda I_{2/(3-Z)} = \int_0^{2/(3-Z)} y^{2n+3}(1 - y)^{n+3}\, dy.$$

(ii) The integration over $OCNP$ is given below.

(16) $F_{1,2}(Z) = [1 - (1 - Z/2)^{2n+4}]/(n + 2)(2n + 4) - 2^{n+2}[(3 - Z)/2]^{(3n+6)}$
$\{B(2n + 4, n + 3) - \lambda I_{(2-Z)/(3-Z)}(2n + 4, n + 3)\}/(n + 2) \qquad (Z \leq 2).$

(iii) In order to integrate over $OCQRA$, we shall integrate over the unit area $OCBA$ and subtract from this the value obtained by integrating over QRB. Thus,

(17) $F_{1,3}(Z) = 1/(n + 2)(2n + 4) - 2^{n+2}[(3 - Z)/2]^{(3n+6)}$
$B(2n + 4, n + 3)/(n + 2).$

Hence the result for $l = 3$ can be expressed as

$$\begin{aligned}
P_r(\theta_1 + \theta_2 + \theta_3 \leq Z) &= \text{const. } \{F_{1,1}(Z) + F_2(Z) + F_3(Z) + F_4(Z)\} \\
&= \text{const. } \{2(n + 2)\{[1 - (1 - Z/2)^{2n+4}]/(n + 2)(2n + 4) \\
&\quad -\lambda \cdot 2^{n+2}[^{(3-Z)}/2]^{3n+6}[I_{2/(3-Z)}(2n + 4, n + 3) \\
&\quad - I_{(2-Z)/(3-Z)}(2n + 4, n + 3)]/(n + 2)\} \\
&\quad + 2(2n + 3)[1 - (1 - Z/2)^{2n+5}]/(2n + 5) - 2[1 - (1 - Z/2)^{2n+4}] \\
&\quad + [1 - (1 - Z)^{n+3}]/(n + 3)\} \qquad (0 \leq Z \leq 1),
\end{aligned}$$

and

$$\begin{aligned}
&= \text{const. } \{F_{1,2}(Z) + F_2(Z) + F_3(Z) + F_4(1)\} \\
&= \text{const. } \Big[2(n + 2)\Big[[1 - (1 - Z/2)^{2n+4}]/(n + 2)(2n + 4) \\
&\quad - 2^{n+2}\left(\frac{3 - Z}{2}\right)^{3n+6}[B(2n + 4, n + 3) - \lambda I_{(2-Z)/(3-Z)}(2n + 4, n + 3)]/(n + 2)\Big] \\
&\quad + 2(2n + 3)[1 - (1 - Z/2)^{2n+5}]/(2n + 5) - 2[1 - (1 - Z/2)^{2n+4}] + 1/n+3\Big] \\
&\qquad (1 \leq Z \leq 2),
\end{aligned}$$

and

$$= \text{const.}\ \{F_{1,3}(Z) + F_2(2) + F_3(2) + F_4(1)\}$$

$$= \text{const.} \left\{ 2(n+2) \left\{ 1/(n+2)(2n+4) - 2^{n+2} \left(\frac{3-Z}{2} \right)^{3n+6} \right.\right.$$

$$\left.\left. B(2n+4, n+3)/(n+2) \right\} + 2(2n+3)/(2n+5) - 2 + 1/(n+3) \right\}$$

$$(2 \leq Z \leq 3),$$

where const. $= (n+2)(n+3)(2n+5)$ and $\lambda = B(2n+4, n+3)$.

The exact distribution is obtained for $l = 4$ by the similar method. The final results are available with the author and are not given here due to lack of space.

The method given in the above sections can be used to find the distribution of the sum of roots of a determinantal equation of any order under the condition $m = 0$.

5. Moments of the distributions. The moments can be obtained by expanding the right-hand side of (6) in terms of x and then collecting the coefficients of x. The moments for $l = 2$ have been derived here and the method is illustrated below:

(a) $l = 2$. Equation (9) gives

$$\int_{0<\theta_2<\theta_1<1} \Pi(1-\theta_i)^n(\theta_1 - \theta_2)e^{x\Sigma\theta_i}\,\Pi\, d\theta_i = \text{const.} \left\{ 2\int_0^1 (1-u)^{2n+2} e^{2xu}\, du \right.$$

$$\left. - \int_0^1 (1-u)^{n+1} e^{xu}\, du \right\} = \text{const.} \left\{ 2\int_0^1 (1-u)^{2n+2} \sum_{t=0}^{\infty} \frac{(2xu)^t}{t!} \right.$$

$$\left. - \int_0^1 (1-u)^{n+1} \sum_{t=0}^{\infty} \frac{(xu)^t}{t!} \right\} = \text{const.} \left\{ 2 \sum_{t=0}^{\infty} \frac{(2x)^t}{t!} \frac{\Gamma(t+1)\Gamma(2n+3)}{\Gamma(2n+t+4)} \right.$$

$$\left. - \sum_{t=0}^{\infty} \frac{x^t}{t!} \frac{\Gamma(t+1)\Gamma(n+2)}{\Gamma(n+t+3)} \right\} = \text{const.} \left\{ \frac{2}{2n+3} \left[1 + \frac{2x}{2n+4} \right.\right.$$

$$\left. + \frac{(2x)^2}{(2n+4)(2n+5)} + \frac{(2x)^3}{(2n+4)(2n+5)(2n+6)} + \cdots \right]$$

$$- \frac{1}{(n+2)} \left[1 + \frac{x}{n+3} + \frac{x^2}{(n+3)(n+4)} \right.$$

$$\left.\left. + \frac{x^3}{(n+3)(n+4)(n+5)} + \cdots \right] \right\}.$$

Thus

$$E(e^{x\Sigma\theta_i}) = \left\{ 1 + \frac{x}{1!} \cdot \frac{3}{(n+3)} + \frac{x^2}{2!} \frac{12(n+2)(4n+11)}{(n+3)(n+4)(2n+4)(2n+5)} \right.$$

$$\left. + \frac{x^3}{3!} \frac{120(n+2)(n+3)(4n+13)}{(2n+4)(2n+5)(2n+6)(n+3)(n+4)(n+5)} + \cdots \right\}.$$

Hence

$$\mu_1' = 3/(n+3),$$
$$\mu_2' = 6(4n+11)/(n+3)(n+4)(2n+5)$$

and

$$\mu_3' = 30(4n+13)/(n+3)(n+4)(n+5)(2n+5).$$

The moments for $l = 3$ and 4 can be obtained in a similar way.

Acknowledgements. The problem was suggested to me by Dr. P. L. Hsu. I take this opportunity to express my gratitude to Dr. P. L. Hsu for guiding me in this research. I am also indebted to Dr. Harold Hotelling for help and suggestions in the work.

REFERENCES

[1] D. N NANDA, "Distribution of a root of a determinantal equation," *Annals of Math Stat.*, Vol. 19 (1948), pp 47–57.

[2] D. N. NANDA, "Limiting distribution of a root of a determinantal equation," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 340–350

NOTES

This section is devoted to brief research and expository articles and other short items.

A NOTE ON THE POWER OF A NON-PARAMETRIC TEST

By F. J. Massey, Jr.

University of Oregon

1. Introduction. Let $x_1 < x_2 < \cdots < x_n$ be the ordered results of n independent observations of a random variable X which has a continuous cumulative distribution function $F(x)$. The following test for the hypothesis that $F(x)$ has some specified form, say $F_0(x)$, has been suggested by Wolfowitz [1].

Form the cumulative distribution of the sample and obtain the maximum deviation of this from $F_0(x)$ Thus if

$$
\begin{aligned}
S_n(x) &= 0 && \text{when } x < x_1, \\
&= \frac{k}{n} && \text{when } x_k \leq x < x_{k+1}, \\
&= 1 && \text{when } x_n < x,
\end{aligned}
$$

the test statistic used would be

$$d = \max_x | F_0(x) - S_n(x) | \sqrt{n},$$

and the hypothesis would be rejected if d is large, say larger than d_α which is so chosen that the probability of a type I error is α. The limiting distribution (as $n \to \infty$) of d has been tabled [2], and a short table of the distribution of d for various small values of n ($n \leq 80$) has been given [3].

The purpose of this note is as follows: 1. A lower bound for the power of the test is given. 2. This test is shown to be consistent against any continuous alternative $F(x) = F_1(x)$, where $F_1(x) \not\equiv F_0(x)$. 3. The test is shown to be biased for finite n. 4. An indication of similar results for a two sample test.

2. Lower bound for the power function. Let $\Delta = \max_x | F_0(x) - F_1(x) |$ and let x_0 be a value of x such that $\Delta = | F_0(x_0) - F_1(x_0) |$. The probability that $d > d_\alpha$ is certainly not less than $\Pr\{\sqrt{n} | F_0(x_0) - S_n(x_0) | > d_\alpha\}$. This is the same as

$$1 - \Pr\left\{F_0(x_0) - \frac{d_\alpha}{\sqrt{n}} < S_n(x_0) < F_0(x_0) + \frac{d_\alpha}{\sqrt{n}}\right\},$$

which, since $S_n(x_0)$ is the proportion of observations falling less or equal to x_0, is given by the binomial probability law.

If $F(x) = F_1(x)$ the probability of an observation being less than x_0 is $F_1(x_0)$. Since $F_0(x_0) = F_1(x_0) \pm \Delta$ the above probability can be written as follows:

$$
\begin{aligned}
&1 - \Pr\{F_1(x_0) \pm \Delta - d_\alpha/\sqrt{n} < S_n(x_0) < F_1(x_0) \pm \Delta + d_\alpha/\sqrt{n}\} \\
&= 1 - \Pr\{\pm\Delta - d_\alpha/\sqrt{n} < S_n(x_0) - F_1(x_0) < \pm \Delta + d_\alpha/\sqrt{n}\} \\
&= 1 - \Pr\{(-d_\alpha \pm \Delta\sqrt{n})/\sqrt{F_1(x_0)(1 - F_1(x_0))} < (S_n(x_0) - F_1(x_0))\sqrt{n}/ \\
&\qquad \sqrt{F_1(x_0)(1 - F_1(x_0))} < (d_\alpha \pm \Delta\sqrt{n})/\sqrt{F_1(x_0)(1 - F_1(x_0))}\}.
\end{aligned}
$$

Δ is fixed. It has been found [3] by observation for samples of size ≤ 80 that d_α actually decreases in size as n increases. For sufficiently large n both

$$-d_\alpha \pm \Delta\sqrt{n} \quad \text{and} \quad d_\alpha \pm \Delta\sqrt{n}$$

have the same sign and the law of large numbers indicates that the above probability approaches zero and the expression approaches unity.

The last expression above can also be used as a lower bound of the power of the test for finite n.

For large values of n this probability is given approximately by the normal distribution Thus we can write for large n;

$$\text{power} > 1 - \int_{\lambda_1}^{\lambda_2} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt,$$

where

$$\lambda_1 = (-d_\alpha \pm \Delta\sqrt{n})/\sqrt{F_1(x_0)(1 - F_1(x_0))}$$

and

$$\lambda_2 = (d_\alpha \pm \Delta\sqrt{n})/\sqrt{F_1(x_0)(1 - F_1(x_0))}.$$

If n is so large that λ_1 and λ_2 are of the same sign and sufficiently different from zero we can replace $F_1(x_0)$ by $\frac{1}{2}$ and not decrease the value of the integral. In this case we might use as a working formula

$$
\begin{aligned}
\lambda_1 &= 2(-d_\alpha \pm \Delta\sqrt{n}), \\
\lambda_2 &= 2(d_\alpha \pm \Delta\sqrt{n}).
\end{aligned}
$$

Since

$$1 - \frac{1}{\sqrt{2\pi}} \int_{\lambda_1}^{\lambda_2} e^{-t^2/2}\, dt$$

approaches one as n tends to infinity, the power, which is larger, must also approach one, and thus the test is consistent.

To demonstrate the biasedness of the test for fixed n consider the following picture.

The $F_0(x)$ is shown as a heavy line and an alternative $F_1(x)$ as a dash-dot line. $F_1(x)$ coincides with $F_0(x)$ except between the point $x = a$ and $x = b$. If $S_n(x)$ falls outside of the indicated band at any point we agree to reject the hypothesis $F(x) = F_0(x)$. If $F(x) = F_1(x)$ the $S_n(x)$ has no chance of being outside the band between $x = a$ and $x = c$, less chance between $x = c$ and $x = b$ than if

$F(x) = F_0(x)$, and the same chance for x larger than b. This indicates that the probability of rejecting $F(x) = F_0(x)$, if actually $F(x) = F_1(x)$, is greater than the probability of rejecting $F(x) = F_0(x)$ if this is actually true. Thus the test is biased.

3. Two sample test. Let $S_n(x)$ and $S'_m(x)$ be the cumulative distributions observed for samples of sizes n and m from two populations having continuous cumulative distribution functions $F(x)$ and $F'(x)$ respectively. Under the assumption that $F(x) = F'(x)$ the limiting distribution (as n and m tend to in-

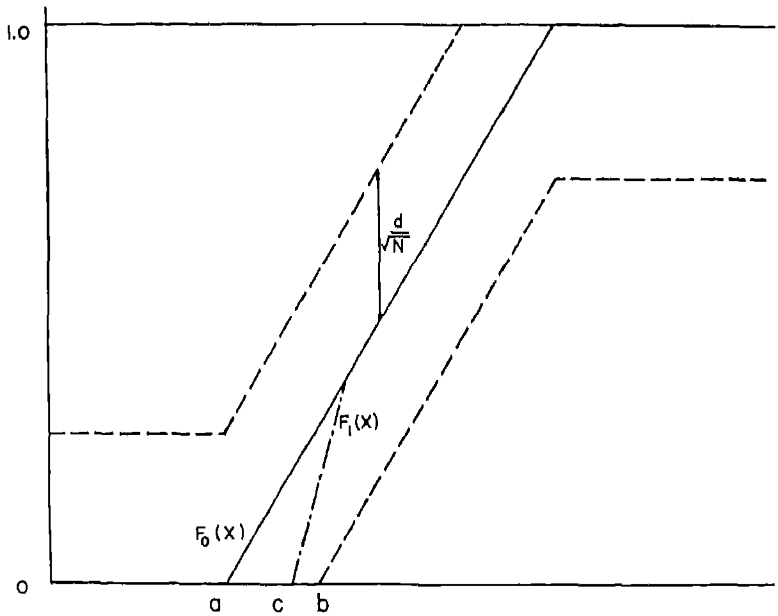


FIG. 1.

finity) of $d' = (n^{-1} + m^{-1})^{-\frac{1}{2}} \max_x | S_n(x) - S'_m(x) |$ has been found and tabled [4], but the distribution of this statistic for small n and m is not known.

Suppose we wish to test the hypothesis that $F(x) = F'(x)$ at level of significance α and agree to reject this if d' is larger than d'_α, where d'_α is the value which would be exceeded a proportion α of the time if the hypothesis is true. The values of d'_α are not known for small samples but are for the limiting case [4].

The same argument as in Section 2 gives a limiting lower bound to the power of the test in terms of

$$\Delta = | F(x_0) - F'(x_0) |,$$

where x_0 is the value of x which maximizes $| F(x) - F'(x) |$, to be

$$1 - \int_{\lambda'_1}^{\lambda'_2} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt,$$

where

$$\lambda_1' = \left(-d_\alpha' \sqrt{\frac{1}{n} + \frac{1}{m}} \pm \Delta\right) \Big/ \sqrt{\frac{F(x_0)[1 - F(x_0)]}{n} + \frac{F'(x_0)[1 - F'(x_0)]}{m}}$$

and

$$\lambda_2' = \left(d_\alpha' \sqrt{\frac{1}{n} + \frac{1}{m}} \pm \Delta\right) \Big/ \sqrt{\frac{F(x_0)[1 - F(x_0)]}{n} + \frac{F'(x_0)[1 - F'(x_0)]}{m}}.$$

Since this lower bound approaches one as n and m approach infinity the power also approaches one and the test is consistent.

REFERENCES

[1] J WOLFWITZ, "Non-parametric statistical inference," *Proceedings of the Symposium on Mathematical Statistics and Probability*, University of California Press, 1949, pp 93–113.
[2] N SMIRNOV, "Table for estimating the goodness of fit of empirical distributions," *Annals of Math Stat.*, Vol 19 (1948), pp 279–281
[3] F MASSEY, "A note on the estimation of a distribution function by confidence limits," *Annals of Math. Stat*, Vol 21 (1950), pp. 116–120.
[4] N SMIRNOV, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bull Math Univ. Moscou, Série Int.*, Vol 2, fasc 2 (1939)

ON OPTIMUM SELECTIONS FROM MULTINORMAL POPULATIONS[1]

BY Z. W. BIRNBAUM AND D. G. CHAPMAN[2]

University of Washington

1. Introduction. Let $Y_1, Y_2, \cdots, Y_n$ be scores in n admission tests such as those used in educational institutions, personnel selection, or testing of materials, and let these scores be used as a basis for selecting a sub-population Π^* from an initial population Π. This selection is usually performed in such a manner that an achievement or performance score X has a distribution in Π^*, which shows some required improvement over the distribution of X in Π; such an improvement may for example consist in changing the expectation $E(X)$ of X in Π to a pre-assigned value $E^*(X)$ in Π^*. Among all selection procedures based on $Y_1, \cdots, Y_n$ and achieving the required improvement of the distribution of X, it appears desirable to find those which retain as large a portion of Π as possible It will be shown that under certain assumptions the linear truncations studied in an earlier paper [1] are such optimal selections.

2. Selection, truncation, linear truncation. Let the frequency of individuals with the scores $(X, Y_1, \cdots, Y_n)$ be $F(X, Y_1, \cdots, Y_n)$ in Π and

[1] Presented at the New York meeting of the Institute of Mathematical Statistics on December 27, 1949

[2] Research done under the sponsorship of the Office of Naval Research.

$$F^*(X, Y_1, \cdots, Y_n)$$

in Π^*. Since Π^* was obtained by selection from Π, we have $F^*/F \leq 1$, and since the selection was made solely on the basis of the values of $Y_1, \cdots, Y_n$, the ratio F^*/F is independent of X. We thus have

$$\frac{F^*(X, Y_1, \cdots, Y_n)}{F(X, Y_1, \cdots, Y_n)} = \varphi(Y_1, \cdots, Y_n)$$

and

$$0 \leqq \varphi(Y_1, \cdots, Y_n) \leq 1. \tag{2.1}$$

Let $N = \iint \cdots \int F(X, Y_1, \cdots, Y_n)\, dX\, dY_1 \cdots dY_n$ and

$$N^* = \iint \cdots \int F^*(X, Y_1, \cdots, Y_n)\, dX\, dY_1 \cdots dY_n$$

be the number of individuals in Π and Π^*, and $f(X, Y_1, \cdots, Y_n)$ and $f^*(X, Y_1, \cdots, Y_n)$ the distribution densities in Π and Π^*, respectively, so that $F = Nf$, $F^* = N^*f^*$ and $\iint \cdots \int f\, dX\, dY_1 \cdots dY_n = \iint \cdots \int f^*\, dX\, dY_1 \cdots dY_n = 1$. We then have

$$N^*f^* = \varphi Nf,$$

and

$$\frac{N^*}{N} = \iint \cdots \int \varphi(Y_1, \cdots, Y_n) f(X, Y_1, \cdots, Y_n)\, dX\, dY_1 \cdots dY_n. \tag{2.2}$$

Thus any selection of a subpopulation Π^* from Π based only on $Y_1, \cdots, Y_n$, defines a $\varphi(Y_1, \cdots, Y_n)$ satisfying (2.1). Conversely, if the frequencies

$$F(X, Y_1, \cdots, Y_n)$$

in Π are given, any measurable $\varphi(Y_1, \cdots, Y_n)$ satisfying (2.1) defines new frequencies $F^* = \varphi F$ and hence a selection from Π based only on $Y_1, \cdots, Y_n$.

These considerations lead to the following definitions:

A measurable function $\varphi(Y_1, \cdots, Y_n)$ which satisfies (2.1) is called a *selection in* $Y_1, \cdots, Y_n$. If, in particular, φ is the characteristic function of a set Ω in $(Y_1, \cdots, Y_n)$, that is $\varphi = 1$ in Ω and $\varphi = 0$ in $\bar{\Omega}$, then the selection φ will be called a *truncation in* $Y_1, \cdots, Y_n$ *to the set* Ω. If Ω is defined by a condition of the form

$$\sum_{j=1}^{n} a_j Y_j \geq t$$

with constant a_j, t, then the truncation to the set Ω will be called a *linear truncation in* $Y_1, \cdots, Y_n$.

In view of (2.2) we will refer to

$$(2.3)\quad r(\varphi) = \int\int \cdots \int \varphi(Y_1, \cdots, Y_n) f(X, Y_1, \cdots, Y_n)\, dX\, dY_1, \cdots dY_n$$

as the *fraction retained in the selection* φ.

3. A lemma. We will need the following slight generalization of the fundamental lemma of Neyman-Pearson (cf. [2]).

LEMMA *Let* $G(Y_1, \cdots, Y_n), G_1(Y_1, \cdots, Y_n), \cdots, G_m(Y_1, \cdots, Y_n)$ *be given integrable functions and* $c_1, \cdots, c_m$ *given constants, and let* (ϕ) *be the family of all measurable functions* $\varphi(Y_1, \cdots, Y_n)$ *which satisfy the conditions*

$$(3.1)\qquad 0 \le \varphi(Y_1, \cdots, Y_n) \le 1$$

$$(3.2)\quad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \varphi(Y_1, \cdots, Y_n) G_i(Y_1, \cdots, Y_n)\, dY_1 \cdots dY_n = c_i \qquad \text{for}\quad i = 1, \cdots, m.$$

If there exist constants $k_1, \cdots, k_m$ *such that the characteristic function* $\varphi_0(Y_1, \cdots, Y_n)$ *of the set* $\underset{(Y_1, \cdots, Y_n)}{E}\left[G \ge \sum_{i=1}^{m} k_i G_i\right] = E$ *belongs to* (ϕ), *then*

$$(3.3)\quad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \varphi_0 G\, dY_1 \cdots dY_n \ge \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \varphi G\, dY_1 \cdots dY_n$$

for any φ *in* (ϕ).

PROOF: We have $\varphi_0 = 1 \ge \varphi$ in E and $\varphi_0 = 0 \le \varphi$ in $\bar{E}$, hence

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(G - \sum_{i=1}^{m} k_i G_i\right) \varphi_0\, dY_1 \cdots dY_n \ge \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(G - \sum_{i=1}^{m} k_i G_i\right) \varphi\, dY_1 \cdots dY_n,$$

and (3.3) follows since φ_0 and φ fulfill (3.2).

4. Selection from a multivariate normal population, for which the fraction retained is maximum. From now on we assume that the conditional distribution of X for given $Y_1, Y_2, \cdots, Y_n$ is normal with a mean which is a linear function of the Y's and with a variance which is independent of them, i.e.,

$$(4.1)\qquad f(X \mid Y_1, Y_2, \cdots, Y_n) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-\left(X - \sum_{i=1}^{n} \rho_i Y_i\right)^2}{2\sigma^2}\right].$$

Let $Q(Y_1, \cdots, Y_n)$ denote the marginal density of $Y_1, \cdots, Y_n$.

THEOREM 1. *A selection such that*

1° *in* Π^* *a proportion at most equal to a given proper fraction* ϵ *has values of* X *below* X_0, *i.e. the* ϵ*-quantile in* Π^* *is greater than or equal to* X_0, *when* X_0 *is a given number greater than the* ϵ*-quantile in* Π,

2° *the fraction retained is maximum,*

is a linear truncation.

PROOF: We have to maximize

$$r(\varphi) = \int \cdots \int \varphi(Y_1, \cdots, Y_n) Q(Y_1, \cdots, Y_n)\, dY_1 \cdots dY_n \tag{4.2}$$

under the condition

$$\frac{\int_{-\infty}^{X_0} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \varphi(Y_1, \cdots, Y_n) Q(Y_1, \cdots, Y_n) f(X \mid Y_1, \cdots, Y_n)\, dY_1 \cdots dY_n dX}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \varphi(Y_1, \cdots, Y_n) Q(Y_1, \cdots, Y_n) f(X \mid Y_1, \cdots, Y_n)\, dY_1 \cdots dY_n dX} \le \epsilon.$$

Substituting the expression (4.1) for $f(X \mid Y_1, \cdots, Y_n)$ and integrating with respect to X we may rewrite this in the form

$$L(\varphi) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \varphi(Y_1, \cdots, Y_n) Q(Y_1, \cdots, Y_n) \cdot \left[\psi\left(\frac{X_0 - \sum_{i=1}^{n} \rho_i Y_i}{\sigma} \right) - \epsilon \right] dY_1 \cdots dY_n \le 0, \tag{4.3}$$

where

$$\psi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2}\, dt,$$

and we have to maximize (4 2) under condition (4.3).

Without loss of generality the inequality $L(\varphi) \le 0$ in (4.3) may be replaced by equality. For if we had a selection φ_1 which maximizes (4.2) and satisfies (4.3) with a strict inequality $L(\varphi_1) < 0$, then φ_1 could not be equal to 1 almost everywhere since then we would have $F^* = F$ almost everywhere and X_0 would be equal to the ϵ-quantile in Π, in contradiction with 1°; hence $\varphi_2 = \varphi_1 + \alpha(1 - \varphi_1)$ for sufficiently small $\alpha > 0$ would also satisfy (4.3) with a strict inequality but would yield $r(\varphi_2) > r(\varphi_1)$.

To solve our problem we now have to maximize (4.2) under the condition

$$L(\varphi) = 0. \tag{4 4}$$

Applying the lemma of Section 3, with $m = 1$, and

$$G(Y_1, \cdots, Y_n) = Q(Y_1, \cdots, Y_n),$$

$$G_1(Y_1, \cdots, Y_n) = Q(Y_1, \cdots, Y_n) \left[\psi\left(\frac{X_0 - \sum_{i=1}^{n} \rho_i Y_i}{\sigma} \right) - \epsilon \right],$$

we conclude that the selection satisfying 1° and 2° will be the characteristic function $\varphi_0(Y_1, \cdots, Y_n)$ of the set defined by

$$k \left[\psi\left(\frac{X_0 - \sum_{i=1}^{n} \rho_i Y_i}{\sigma} \right) - \epsilon \right] \le 1, \tag{4.5}$$

provided k can be determined so that φ_0 satisfies (4.4).

To find such a k we consider

$$I(t) = \int_{\sum_{i=1}^{n} \rho_i Y_i \geq t} \cdots \int Q(Y_1, \cdots, Y_n) \left[\psi\left(\frac{X_0 - \sum_{i=1}^{n} \rho_i Y_i}{\sigma}\right) - \epsilon \right] dY_1 \cdots dY_n .$$

As t tends to $-\infty$, $I(t)$ tends to $L(1)$, where L was defined by (4.3). Since the ϵ-quantile in Π was less than X_0 it follows that $I(-\infty) = L(1) > 0$. Since $I(t) < 0$ for large t, there exists t_0 such that $I(t_0) = 0$, and clearly,

$$\psi\left(\frac{X_0 - t_0}{\sigma}\right) - \epsilon > 0.$$

Setting in (4.5) $k = [\psi((X_0 - t_0)/\sigma) - \epsilon]^{-1}$, one obtains a φ_0 such that

$$L(\varphi_0) = I(t_0) = 0.$$

The selection φ_0 is the linear truncation to the set $\sum_{i=1}^{n} \rho_i Y_i \geq t_0$.

By a similar and somewhat simpler argument one proves the following theorem.

THEOREM 2 *A selection such that*

1° *in* Π^* *the mean of* X *has a value greater than or equal to a pre-assigned number* $m > 0$,

2° *the fraction retained is maximum,*

is a linear truncation to a set $\sum_{i=1}^{n} \rho_i Y_i \geq t_0$.

An immediate consequence of Theorems 1 and 2 is that a linear truncation, using a properly determined weighted score $\sum_{i=1}^{n} \rho_i Y_i$ and cutting score t_0, is more economical than any truncation to a set $Y_i \geq t_i$, $i = 1, 2, \cdots, n$, that is than any truncation performed on each admission score separately.

REFERENCES

[1] Z W. BIRNBAUM, "Effect of linear truncation on a multinormal population," *Annals of Math Stat*, Vol. 21 (1950), pp. 272–279.

[2] J NEYMAN AND E S PEARSON, "Contributions to the theory of testing statistical hypotheses," *Stat. Res Memoirs*, Vol I (1936), pp. 1–37, particularly pp 10–11.

THE DISTRIBUTION OF DISTANCE IN A HYPERSPHERE

BY J. M HAMMERSLEY

University of Oxford

1. Summary. Deltheil ([1], pp. 114–120) has considered the distribution of distance in an n-dimensional hypersphere. In this paper I put his results (17) in a more compact form (16); and I investigate in greater detail the asymptotic form of the distribution for large n, for which the rather surprising result emerges that this distance is almost always nearly equal to the distance between the

extremities of two orthogonal radii. I came to study this distribution by the need to compute a doubly-threefold integral, which measures the damage caused to plants by the presence of radioactive tracers in their fertilizers; for the distribution affords a method of evaluating numerically certain multiple integrals. I hope to describe elsewhere this application of the theory.

2. Derivation of the frequency function. Let T_1 and T_2 be vector spaces of n and $2n$ dimensions respectively. Let P and Q be any pair of points in T_1. Denote by (PQ) the point in T_2, whose first n coordinates are the coordinates of P in T_1 and whose last n coordinates are the coordinates of Q in T_1. Let $\{P\}$ and $\{Q\}$ be point sets in T_1, and let $\{PQ\}$ be the point set in T_2 such that $(PQ) \in \{PQ\}$ if and only if both $P \in \{P\}$ and $Q \in \{Q\}$. Let $M_1\{P\}$ denote the n-dimensional measure of the point set $\{P\}$ in T_1, and let $M_2\{PQ\}$ denote the $2n$-dimensional measure of the point set $\{PQ\}$ in T_2. Then

$$M_2\{PQ\} = \int_{\{P\}} M_1\{Q\}\, dM_1\{P\}. \tag{1}$$

Let R be a fixed point in T_1, and let $S_n(a)$ be the n-dimensional hypersphere in T_1 with centre R and radius a. Let A and B be any two points chosen at random in $S_n(a)$, the distributions of A and B being independent and uniform over the interior of $S_n(a)$. Denote the distance AB by r, and let $\lambda = r/2a$, so that λ may take any value in the interval $0 \le \lambda \le 1$. We require the frequency function of λ, which we shall denote by $f_n(\lambda)$.

The volume content of $S_n(a)$ is

$$V_n(a) = \pi^{n/2}a^n/\Gamma(\tfrac{1}{2}n + 1); \tag{2}$$

and the content of the segment of the surface of $S_n(a)$ bounded by a right hyperspherical cone, whose vertex is at R and whose line generators make a fixed semi-vertical angle θ with a fixed radius of $S_n(a)$, is

$$U_n(a, \theta) = \frac{2\pi^{(n-1)/2}a^{n-1}}{\Gamma(\tfrac{1}{2}n - \tfrac{1}{2})} \int_0^\theta \sin^{n-2}\phi\, d\phi. \tag{3}$$

As a particular case of (2), the whole surface of $S_n(a)$ has content

$$U_n(a, \pi) = 2\pi^{n/2}a^{n-1}/\Gamma(\tfrac{1}{2}n). \tag{4}$$

Let $\{AB\}$ be the point set in T_2 such that $(AB) \in \{AB\}$ if and only if the corresponding points A and B satisfy all the inequalities

$$0 \le RA \le a, \qquad 0 \le RB \le a, \qquad r \le AB \le r + dr. \tag{5}$$

Then, by the definition of $f_n(\lambda)$,

$$M_2\{AB\} \propto f_n(r/2a)\, dr;$$

but since

$$\int_0^{2a} M_2\{AB\}\, dr = V_n^2, \qquad \int_0^{2a} f_n(r/2a)\, dr/2a = 1,$$

we have

$$M_2\{AB\} = V_n^2 f_n(r/2a)\,dr/2a \equiv p_n(r, a)\,dr, \qquad \text{say.} \tag{6}$$

Consider also the point set $\{CD\}$ in T_2 such that $(CD) \in \{CD\}$ if and only if the corresponding points C and D satisfy all the inequalities

$$0 \leq RC \leq a + da, \qquad a \leq RD \leq a + da, \qquad r \leq CD \leq r + dr. \tag{7}$$

For each fixed D of $\{D\}$, C is constrained to lie on the segment of the hyperspherical shell of thickness dr, radius r, and centre D, bounded by the intersection of this shell with $S_n(a + da)$ The hyperspherical cone, with vertex D, whose line generators all pass through this intersection, has a semi-vertical angle θ given by

$$\cos\theta = r/2a = \lambda, \tag{8}$$

and so, from (3), the M_1 of all C which satisfy (7) for each fixed D is $U_n(r, \arccos \lambda)\,dr$ On the other hand the M_1 of all D which satisfy (7) is the content of the hyperspherical shell of thickness da, radius a, and centre R, and is thus $U_n(a, \pi)\,da$ by virtue of (4) Consequently, from (1)

$$M_2\{CD\} = U_n(r, \arccos\lambda)U_n(a, \pi)\,da\,dr. \tag{9}$$

On the other hand, by symmetry, $M_2\{CD\} = \frac{1}{2}M_2\{EF\}$, where $(EF) \in \{EF\}$ if and only if the corresponding points E and F satisfy either all the inequalities

$$0 \leq RE \leq a + da, \qquad a \leq RF \leq a + da, \qquad r \leq EF \leq r + dr,$$

or all the inequalities

$$0 \leq RF \leq a + da, \qquad a \leq RE \leq a + da, \qquad r \leq EF \leq r + dr.$$

We can express this in another way by saying that $(EF) \in \{EF\}$ if and only if the corresponding points E and F satisfy all the inequalities

$$0 \leq RE \leq a + da, \qquad 0 \leq RF \leq a + da, \qquad r \leq EF \leq r + dr,$$

but do not satisfy all the inequalities

$$0 \leq RE \leq a, \qquad 0 \leq RF \leq a, \qquad r \leq EF \leq r + dr.$$

From this second point of view we see that

$$M_2\{EF\} = p_n(r, a + da)\,dr - p_n(r, a)\,dr = \frac{\partial}{\partial a}p_n(r, a)\,dr\,da;$$

and so

$$M_2\{CD\} = \tfrac{1}{2}\frac{\partial}{\partial a}p_n(r, a)\,dr\,da. \tag{10}$$

Then from (2), (3), (4), (6), (9), and (10).

$$\frac{1}{2}\frac{\partial}{\partial a}\left\{\frac{\pi^n a^{2n}}{[\Gamma(\frac{1}{2}n + 1)]^2} f_n\left(\frac{r}{2a}\right)\cdot\frac{1}{2a}\right\} = \left\{\frac{2\pi^{(n-1)/2} r^{n-1}}{\Gamma(\frac{1}{2}n - \frac{1}{2})}\int_0^{\arccos\lambda} \sin^{n-2}\phi\,d\phi\right\}\left\{\frac{2\pi^{n/2} a^{n-1}}{\Gamma(\frac{1}{2}n)}\right\}. \tag{11}$$

By performing the partial differentiation on the left-hand side, then substituting $z = \cos\phi$ and $r = 2a\lambda$, and using the relations

$$\Gamma(\tfrac{1}{2}n + 1) = \tfrac{1}{2}n\Gamma(\tfrac{1}{2}n), \qquad \pi^{1/2}\Gamma(n+1) = 2^n\Gamma(\tfrac{1}{2}n + \tfrac{1}{2})\Gamma(\tfrac{1}{2}n + 1),$$
$$B(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}n + \tfrac{1}{2}) = \{\Gamma(\tfrac{1}{2}n + \tfrac{1}{2})\}^2/\Gamma(n+1),$$

we reduce (11) to the form

$$(12)\quad (2n-1)f_n(\lambda) - \lambda f_n'(\lambda) = \frac{2n(n-1)\lambda^{n-1}}{B(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}n + \tfrac{1}{2})}\int_\lambda^1 (1 - z^2)^{(n-3)/2}\,dz.$$

We multiply (12) by $-\lambda^{-2n}$ and use the reduction formula

$$(13)\quad (n-1)\int_\lambda^1 (1-z^2)^{(n-3)/2}\,dz = n\int_\lambda^1 (1-z^2)^{(n-1)/2}\,dz + \lambda(1-\lambda^2)^{(n-1)/2}.$$

Each side of the resulting equation is a perfect differential coefficient, and upon integration we obtain

$$(14)\qquad f_n(\lambda) = \frac{2n\lambda^{n-1}}{B(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}n + \tfrac{1}{2})}\int_\lambda^1 (1-z^2)^{(n-1)/2}\,dz + c\lambda^{2n-1},$$

where C is the constant of integration. We obtain the cumulative distribution function by integrating (14) over 0 to λ,

$$(15)\quad F_n(\lambda) = (2\lambda)^n I_{1-\lambda^2}(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}) + I_{\lambda^2}(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}n + \tfrac{1}{2}) + C\lambda^{2n}/2n,$$

where $I_x(p, q)$ is the incomplete beta-function ratio

$$I_x(p, q) = \int_0^x z^{p-1}(1-z)^{q-1}\,dz/B(p, q)$$

tabulated by Pearson [2]. Putting $\lambda = 1$ in (15) we get

$$1 = F_n(1) = 1 + C/2n;$$

so $C = 0$, and we have the final result

$$(16)\qquad f_n(\lambda) = 2^n n\lambda^{n-1} I_{1-\lambda^2}(\tfrac{1}{2}n + \tfrac{1}{2}, \tfrac{1}{2}).$$

This compact form may be compared with Deltheil's expression [1] for the frequency function of r, namely

$$(17)\qquad g_n(r) = \frac{n^2 r^{n-1}}{a^{2n}}\int_{r/2}^a \rho^{n-1} h_n\left(\frac{r}{\rho}\right)d\rho,$$

where

$$h_n(2\sin\theta) = \int_0^{\frac{1}{2}\pi - \theta}\sin^{n-2}\phi\,d\phi \Big/ \int_0^{\frac{1}{2}\pi}\sin^{n-2}\phi\,d\phi,$$

expressions which he evaluates only for the particular cases $n = 3, 5, 7, 9$.

Interesting particular cases of (16) are

$$f_1(\lambda) = 2(1 - \lambda), \qquad f_2(\lambda) = \frac{16}{\pi}\lambda\{\arccos\lambda - \lambda(1-\lambda^2)^{1/2}\}, \tag{18}$$
$$f_3(\lambda) = 12\lambda^2(1-\lambda)^2(2+\lambda),$$

which give the appropriate frequency functions for a line, a circle, and a sphere respectively.

3. Recurrence relations and moments of the distribution. From (13) and (14) we have a recurrence relation for penadjacent values of n,

$$\frac{f_n(\lambda)}{n} = 4\lambda^2\frac{f_{n-2}(\lambda)}{n-2} - \frac{2\Gamma(n)}{\{\Gamma(\frac{1}{2}n+\frac{1}{2})\}^2}\lambda^n(1-\lambda^2)^{(n-1)/2}. \tag{19}$$

In connection with (18) this shows that

$$f_{2n+1}(\lambda) = P_{4n+1}(\lambda), \qquad f_{2n}(\lambda) = P_{2n-1}(\lambda)\arccos\lambda + P_{4n-2}(\lambda)(1-\lambda^2)^{1/2}, \tag{20}$$

where $P_N(\lambda)$ denotes an unspecified polynomial in λ of degree N or less.

From (16) the rth moment of $f_n(\lambda)$ about $\lambda = 0$ is

$$\mu'_{nr} = \left\{\frac{n\Gamma(n+1)}{\Gamma(\frac{1}{2}n+\frac{1}{2})}\right\}\left\{\frac{\Gamma(\frac{1}{2}n+\frac{1}{2}r+\frac{1}{2})}{(n+r)\,\Gamma(n+\frac{1}{2}r+1)}\right\}. \tag{21}$$

I have not been able to obtain the characteristic function of $f_n(\lambda)$ explicitly from (21) it appears to be of a higher type than the hypergeometric function.

4. The asymptotic form of the distribution for large n. The distribution function is, by (15),

$$F_n(\lambda) = (2\lambda)^n I_{1-\lambda^2}(\tfrac{1}{2}n+\tfrac{1}{2}, \tfrac{1}{2}) + I_{\lambda^2}(\tfrac{1}{2}n+\tfrac{1}{2}, \tfrac{1}{2}n+\tfrac{1}{2}). \tag{22}$$

We show firstly that as $n \to \infty$ the first term of this expression tends to zero. This term is clearly zero if $\lambda = 0$. If $\lambda > 0$

$$\int_0^{1-\lambda^2} z^{(n-1)/2}(1-z)^{-1/2}\,dz \le \lambda^{-1}\int_0^{1-\lambda^2} z^{(n-1)/2}\,dz = (1-\lambda^2)^{(n+1)/2}/\tfrac{1}{2}(n+1)\,\lambda.$$

Hence

$$(2\lambda)^n I_{1-\lambda^2}(\tfrac{1}{2}n+\tfrac{1}{2}, \tfrac{1}{2}) \le \frac{(2\lambda)^n\Gamma(\frac{1}{2}n+1)}{\pi^{1/2}\Gamma(\frac{1}{2}n+\frac{1}{2})}\cdot\frac{(1-\lambda^2)^{(n+1)/2}}{(\frac{1}{2}n+\frac{1}{2})\lambda}$$
$$\le \frac{2\Gamma(\frac{1}{2}n+1)}{\pi^{1/2}\Gamma(\frac{1}{2}n+\frac{3}{2})}(1-\lambda^2)\{4\lambda^2(1-\lambda^2)\}^{(n-1)/2} \le \frac{2\Gamma(\frac{1}{2}n+1)}{\pi^{1/2}\Gamma(\frac{1}{2}n+\frac{3}{2})} \to 0$$

as $n \to \infty$. Secondly, as $n \to \infty$

$$I_{\lambda^2}(\tfrac{1}{2}n+\tfrac{1}{2}, \tfrac{1}{2}n+\tfrac{1}{2}) \sim N_{\lambda^2}(\tfrac{1}{2}, 1/4(n+1)) \sim N_{\lambda^2}(\tfrac{1}{2}, 1/4n),$$

(see Cramér [3] p. 252 with $p = q = \frac{1}{2}$), where $N_x(\mu, \sigma^2)$ is the normal cumulative distribution function of x for mean μ and variance σ^2. Hence λ is asymptotically distributed as $N_\lambda(1/\sqrt{2}, 1/8n)$; and the asymptotic distribution of r is $N_r(a\sqrt{2}, a^2/2n)$. This establishes the result stated in the summary.

It can also be proved, by considering the limiting form of the recurrence relation (19), that the frequency function f_n is asymptotically normal. The main difficulty of proving this fact lies in showing that the frequency function actually possesses a limiting form; and the proof is rather too long to be given here.

REFERENCES

[1] R. DELTHEIL, *Probabilités Géométriques. Traité du Calcul des Probabilités et de ses Applications: Tome II, Fascicule II*, Gauthier-Villars, 1926.
[2] K. PEARSON, *Tables of the Incomplete Beta-Function*, Cambridge University Press, 1934.
[3] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

A NOTE ON THE ASYMPTOTIC SIMULTANEOUS DISTRIBUTION OF THE SAMPLE MEDIAN AND THE MEAN DEVIATION FROM THE SAMPLE MEDIAN

BY R. K. ZEIGLER

Bradley University

Consider a random sample of $2k + 1$ values from a one-dimensional distribution of the continuous type with cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x) = F'(x)$. Let the mean, standard deviation and median of the distribution be denoted by m, σ and θ respectively (θ assumed to be unique). We shall suppose that in some neighborhood of $x = \theta$, $f(x)$ has a continuous derivative $f'(x)$.

If we arrange the sample values in ascending order of magnitude:

$$x_1 < x_2 < \cdots < x_{2k+1},$$

there is a unique sample median x_{k+1} which we shall denote by ξ. The mean deviation from the sample median is then defined by

$$M = \frac{1}{2k} \sum_{i=1}^{2k+1} | x_i - \xi |.$$

In the material that follows we shall assume that the sample items have been ordered only to the extent that k of them are less than ξ and k of them are greater than ξ.

We then have the following

THEOREM. *Let $f(x)$ be a pdf with finite second moment, continuous at $x = \theta$ with $f(\theta) \neq 0$. Then the simultaneous distribution of ξ and M is asymptotically normal. The means of the limiting distribution are θ, the population median, and u', the mean deviation from the population median, while the asymptotic variances are $1/4f^2(\theta)2k$ and $((m - \theta)^2 + \sigma^2 - u'^2)/2k$. The asymptotic expression for the correlation coefficient is $(m - \theta)/\sqrt{(m - \theta)^2 + \sigma^2 - u'^2}$.*

PROOF: Let $u = (M - u')\sqrt{2k}$ and $v = (\xi - \theta)\sqrt{2k}$, where $u' = E | x - \theta |$. Then the simultaneous characteristic function of the two random variables u

and v is given by the following:

$$\begin{aligned}
\phi(t_1, t_2) &= E[e^{it_1u+it_2v}] \\
&= E[e^{it_1(M-u')\sqrt{2k}+it_2(\xi-\theta)\sqrt{2k}}] \\
&= E\exp\left[it_1\left(\frac{1}{2k}\sum_{i=1}^{2k+1}|x_i-\xi|-u'\right)\sqrt{2k}+it_2(\xi-\theta)\sqrt{2k}\right] \\
&= \frac{(2k+1)!}{(k!)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\xi}\cdots\int_{-\infty}^{\xi}\int_{\xi}^{\infty}\cdots\int_{\xi}^{\infty} \\
&\qquad\cdot\exp\left[it_1\left\{\frac{\sum_{i=k+2}^{2k+1}x_i-\sum_{j=1}^{k}x_j}{2k}-u'\right\}\sqrt{2k}+it_2(\xi-\theta)\sqrt{2k}\right] \\
&\qquad f(x_1)f(x_2)\cdots f(x_k)f(x_{k+2})\cdots f(x_{2k+1})f(\xi) \\
&\qquad\quad dx_{2k+1}\cdots dx_{k+2}dx_k\cdots dx_1 d\xi \\
&= \frac{(2k+1)!}{(k!)^2}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\xi}\exp\left\{-\frac{it_1}{\sqrt{2k}}(x+u')\right\}f(x)\,dx\right]^k \\
&\qquad\left[\int_{\xi}^{\infty}\exp\left\{\frac{it_1}{\sqrt{2k}}(x-u')\right\}f(x)\,dx\right]^k e^{it_2(\xi-\theta)\sqrt{2k}}f(\xi)\,d\xi.
\end{aligned}$$

Upon making the substitution $\xi = \theta + y/\sqrt{2k}$, the above expression can be reduced to the following form:

$$\begin{aligned}
(1)\quad \phi(t_1, t_2) = \frac{(2k+1)!}{\sqrt{2k}(k!)^2}\int_{-\infty}^{\infty}&\left\{\left[\int_{-\infty}^{\theta}\exp\left[-\frac{it_1}{\sqrt{2k}}(x+u')\right]f(x)\,dx\right.\right. \\
&\left.+\int_{\theta}^{\theta+(y/\sqrt{2k})}\exp\left[-\frac{it_1}{\sqrt{2k}}(x+u')\right]f(x)\,dx\right] \\
&\cdot\left[\int_{\theta}^{\infty}\exp\left[\frac{it_1}{\sqrt{2k}}(x-u')\right]f(x)\,dx\right. \\
&\left.\left.-\int_{\theta}^{\theta+(y/\sqrt{2k})}\exp\left[\frac{it_1}{\sqrt{2k}}(x-u')\right]f(x)\,dx\right]\right\}^k \\
&\cdot e^{it_2y}f\left(\theta+\frac{y}{\sqrt{2k}}\right)dy.
\end{aligned}$$

Now

$$\int_{-\infty}^{\theta}\exp\left[-\frac{it_1}{\sqrt{2k}}(x+u')\right]f(x)\,dx = \tfrac{1}{2}-\frac{it_1}{\sqrt{2k}}\int_{-\infty}^{\theta}(x+u')f(x)\,dx - \frac{t_1^2}{2(2k)}\int_{-\infty}^{\theta}(x+u')^2f(x)\,dx+\frac{\zeta_1(2k, t_1)}{2k};$$

and

$$\int_\theta^\infty \exp\left[\frac{it_1}{\sqrt{2k}}(x - u')\right] f(x)\,dx = \tfrac{1}{2} + \frac{it_1}{\sqrt{2k}}\int_\theta^\infty (x - u')f(x)\,dx - \frac{t_1^2}{2(2k)}\int_\theta^\infty (x - u')^2 f(x)\,dx + \frac{\zeta_2(2k, t_1)}{2k},$$

where for every fixed t_1, $\zeta_1(2k, t_1)$ and $\zeta_2(2k, t_1) \to 0$ as $k \to \infty$. Similarly, under the substitution $x = (z/\sqrt{2k}) + \theta$,

$$\int_\theta^{\theta+(y/\sqrt{2k})} \exp\left[\frac{it_1}{\sqrt{2k}}(x - u')\right] f(x)\,dx = \frac{1}{\sqrt{2k}}\int_0^y f\left(\frac{z}{\sqrt{2k}} + \theta\right) dz + \frac{it_1}{2k}\int_0^y \left(\frac{z}{\sqrt{2k}} + \theta - u'\right) f\left(\frac{z}{\sqrt{2k}} + \theta\right) dz + \frac{\zeta_3(2k, t_1)}{2k},$$

and

$$\int_\theta^{\theta+(y/\sqrt{2k})} \exp\left[-\frac{it_1}{\sqrt{2k}}(x + u')\right] f(x)\,dx = \frac{1}{\sqrt{2k}}\int_0^y f\left(\frac{z}{\sqrt{2k}} + \theta\right) dz - \frac{it_1}{2k}\int_0^y \left(\frac{z}{\sqrt{2k}} + \theta + u'\right) f\left(\frac{z}{\sqrt{2k}} + \theta\right) dz + \frac{\zeta_4(2k, t_1)}{2k},$$

where $\zeta_3(2k, t_1)$ and $\zeta_4(2k, t_1) \to 0$ as $k \to \infty$ for each fixed t_1. Substituting these expressions in (1) and performing the indicated multiplications we find after some calculation that (1) can be reduced to the following form:

$$\phi(t_1, t_2) = \int_{-\infty}^\infty \frac{(2k + 1)!}{\sqrt{2k}(k!)^2 2^{2k}} \left[1 - \frac{t_1^2(\sigma^2 - u'^2) - 4it_1(m - \theta)yf\left(\frac{z_1}{\sqrt{2k}} + \theta\right)}{2k} + \frac{-4\left\{yf\left(\frac{z_1}{\sqrt{2k}} + \theta\right)\right\}^2 + \zeta(2k, t_1)}{2k}\right]^k e^{it_2 y} f\left(\theta + \frac{y}{\sqrt{2k}}\right) dy,$$

where $0 < z_1 < y$ and $\zeta(2k, t_1) \to 0$ for every fixed t_1 as $k \to \infty$. Now taking the limit as $k \to \infty$, we have

$$\lim_{k\to\infty} \phi(t_1, t_2) = \int_{-\infty}^\infty \sqrt{\frac{2}{\pi}} \exp\left[-\frac{t_1^2}{2}(\sigma^2 - u'^2) + \frac{4it_1(m - \theta)f(\theta)y}{2} - \frac{4[f^2(\theta)]y^2}{2} + it_2 y\right] f(\theta)\,dy.$$

Upon performing the integration,

$$\lim_{k\to\infty} \phi(t_1, t_2) = \exp\left[-\tfrac{1}{2}\left\{t_1^2[(m - \theta)^2 + \sigma^2 - u'^2] + \frac{2t_1 t_2(m - \theta)}{2f(\theta)} + \frac{t_2^2}{4f^2(\theta)}\right\}\right].$$

Since $\sigma^2 > u'^2$, this is the characteristic function for two variables which are normally distributed Thus, the simultaneous distribution of ξ and M is asymptotically normal. It is of interest to note that, if the pdf $f(x)$ is symmetric, the correlation coefficient is zero, and M and ξ are asymptotically independent. We might also note that $\phi(t_1, 0)$ is the characteristic function for the mean deviation from the sample median. Thus, the random variable M is asymptotically normal with asymptotic mean and variance u' and $((m - \theta)^2 + \sigma^2 - u'^2)/2k$ respectively.

The author wishes to express his appreciation to Professor A. T. Craig for valuable suggestions in the study of this problem.

REFERENCES

[1] H CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946
[2] R. K ZEIGLER, "On the mean deviation from the median," unpublished thesis, State University of Iowa

NOTE ON THE EXTENSION OF CRAIG'S THEOREM TO NON-CENTRAL VARIATES

BY OSMER CARPENTER

Carbide and Carbon Chemical Corporation, Oak Ridge

A theorem due to A. T. Craig [1] and H Hotelling [3] concerning the distribution of real quadratic forms in normal variates is extended to the case of non-central normal variates with equal variance.

The following notation is used: A, A_1, A_2 are real symmetric matrices, L is an orthogonal matrix, Γ is a diagonal matrix of latent roots, and X, Y, M and U are column vectors.

THEOREM. *Let $X' = (x_1, \cdots, x_n)$ be a set of normally and independently distributed variates with equal variance σ^2 and means $M' = (m_1, \cdot \cdot, m_n)$. Then, a necessary and sufficient condition that a real symmetric quadratic form $Q(X) = X'AX$ of rank r be distributed as $\sigma^2\chi^2$, where*

$$p(\chi^2, r, \lambda^2) = \tfrac{1}{2}e^{-\lambda^2}(\chi^2/2)^{(r-2)/2}\, e^{-\chi^2/2} \tag{1}$$
$$\sum_{j=0}^{\infty} (\lambda^2\chi^2/2)^j/j!\,\Gamma[(r - 2j)/2],$$

is that $A^2 = A$. If $Q(X)/\sigma^2$ is distributed by $p(\chi^2, r, \lambda^2)$, then $\lambda^2 = Q(M)/2\sigma^2$.

Further, let $Q_1(X) = X'A_1X$ and $Q_2(X) = X'A_2X$ be real symmetric quadratic forms of ranks r_1 and r_2. Then a necessary and sufficient condition that $Q_1(X)$ and $Q_2(X)$ be statistically independent is that $A_1A_2 = 0$.

PROOF. The theorem is proved by establishing the equivalence and factorization of moment generating functions [4]. The moment generating function of

$p(\chi^2, r, \lambda^2)$ is

$$G(t) = Ee^{t\chi^2/2} = e^{\lambda^2 t/(1-t)}(1-t)^{-r/2} \tag{2}$$

Let $x_1, \cdots, x_n$ be normally and independently distributed with means $E(x_i) = m_i$ and common variance σ^2. Without loss of generality, we may take $\sigma^2 = 1$, changing to the general case when necessary with the transformation $x_i = z_i/\sigma$.

Let $Q(X) = X'AX$ be a real symmetric quadratic form of rank r. Then the moment generating function of $Q(X)$ is

$$G_Q(t) = Ee^{tQ(X)/2} = (2\pi)^{-n/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(X-M)'(X-M)-X'tAX]} \prod_1^n dx_i\,. \tag{3}$$

If t is restricted to values such that $|t| < |1/\gamma_0|$, where γ_0 is the dominant latent root of A, then $I - tA$ is positive definite and

$$\begin{aligned} G_Q(t) &= (2\pi)^{-n/2} e^{\frac{1}{2}M'tA(I-tA)^{-1}M} \\ &\quad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}[X-(I-tA)^{-1}M]'(I-tA)[X-(I-tA)^{-1}M]} \prod_1^n dx_i \\ &= e^{\frac{1}{2}M'tA(I-tA)^{-1}M}\, |I - tA|^{-\frac{1}{2}}. \end{aligned} \tag{4}$$

If L is an orthogonal matrix such that

$$L'AL = \Gamma = \begin{pmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \gamma_n \end{pmatrix},$$

where the γ_i are the latent roots of A, then the transformation $M = LU$ gives

$$G_Q(t) = e^{\frac{1}{2}U't\Gamma(I-t\Gamma)^{-1}U}\, |I - t\Gamma|^{-\frac{1}{2}}. \tag{5}$$

A necessary and sufficient condition that $G_Q(t) = G(t)$ is that $A^2 = A$. If $A^2 = A$, then all of the latent roots of A are $+1$ or 0, and sufficiency can be established by substituting the appropriate value of each γ_i into equation (5), giving

$$G_Q(t) = e^{\lambda^2 t/(1-t)}(1-t)^{-r/2} = G(t). \tag{6}$$

Also $\lambda^2 = \sum_1^n \gamma_i u_i^2/2 = \frac{1}{2}(U'\Gamma U) = \frac{1}{2}(M'AM) = Q(M)/2$.

It is apparent from the form of $G_Q(t)$ that a necessary condition for $G_Q(t) = G(t)$ is that $|I - tA|^{-\frac{1}{2}} = (1-t)^{-r/2}$. But it has been proved by Craig [1] that the condition $A^2 = A$ is necessary, as well as sufficient, for this equality.

Next, let $Q_1(X) = X'A_1X$ and $Q_2(X) = X'A_2X$ be real symmetric quadratic forms of ranks r_1 and r_2. Then from (4)

$$\begin{aligned} G(t_1, t_2) &= Ee^{t_1Q_1/2+t_2Q_2/2} \\ &= e^{\frac{1}{2}M'(t_1A_1+t_2A_2)(I-t_1A_1-t_2A_2)^{-1}M}\, |I - t_1A_1 - t_2A_2|^{-\frac{1}{2}}, \end{aligned} \tag{7}$$

t_1, t_2 being restricted to values for which $(I - t_1A_1 - t_2A_2)$ is positive definite. A necessary and sufficient condition that $G(t_1, t_2) = G_Q(t_1)\, G_Q(t_2)$ is $A_1A_2 = 0$ The required equation in the moment generating functions is

$$G(t_1, t_2) = e^{\frac{1}{2}M' t_1 A_1 (I - t_1 A_1)^{-1} M} \, | I - t_1 A_1 |^{-\frac{1}{2}} \, e^{\frac{1}{2}M' t_2 A_2 (I - t_2 A_2)^{-1} M} \, | I - t_2 A_2 |^{-\frac{1}{2}} \tag{8}$$

Assume $A_1A_2 = 0$. Then $(I - t_1A_1 - t_2A_2) = (I - t_1A_1)(I - t_2A_2)$ and $| I - t_1A_1 - t_2A_2 | = | I - t_1A_1 | \cdot | I - t_2A_2 |$. Also

$$(t_1A_1 + t_2A_2)(I - t_1A_1 - t_2A_2)^{-1} = t_1A_1(I - t_1A_1)^{-1} + t_2A_2(I - t_2A_2)^{-1},$$

for using the identity $tA(I - tA)^{-1} = (I - tA)^{-1} - I$, this becomes

$$(I - t_2A_2)^{-1}(I - t_1A_1)^{-1} = (I - t_1A_1)^{-1} + (I - t_2A_2)^{-1} - I$$

Multiplying both sides on the left by $(I - t_2A_2)$ and on the right by $(I - t_1A_1)$, the identity follows. Thus the condition is sufficient.

It is apparent from the form of the moment generating functions that a necessary condition for $G(t_1, t_2) = G_Q(t_1)G_Q(t_2)$ is that $| I - t_1A_1 - t_2A_2 | = | I - t_1A_1 | \; | I - t_2A_2 |$. However, it has been proved by Hotelling [3] and Craig [2] that the condition $A_1A_2 = 0$ is necessary for this equality.

An extension can be made to correlated variates. Let $X' = (x_1, \cdots, x_n)$ be normally distributed with non-singular correlation matrix B and means $M' = (m_1, \cdots, m_n)$. Then there exists a non-singular transformation $X = TZ$, such that the variates Z are independent and have unit variance. Thus $T^{-1}BT'^{-1} = I$, $B = TT'$ and $Q(X) = X'AX = Z'T'ATZ$. Applying the theorem proved above, a necessary and sufficient condition that $Q(X)$ be distributed as χ^2 is that $(T'AT)^2 = T'ABAT = T'AT$, or that $ABA = A$. As before, $\lambda^2 = Q(M)/2$. In the same manner, a necessary and sufficient condition for independence of $Q_1(X)$ and $Q_2(X)$ is that $(T'A_1T)(T'A_2T) = T'A_1BA_2T = 0$, or that $A_1BA_2 = 0$.

REFERENCES

[1] ALLEN T CRAIG, "A note on the independence of certain quadratic forms," *Annals of Math Stat*, Vol 14 (1943), page 195

[2] ALLEN T. CRAIG, "Bilinear forms in normally correlated variables," *Annals of Math. Stat*, Vol 18 (1947), page 565

[3] H. HOTELLING, "A note on a matric theorem of A T Craig," *Annals of Math. Stat*, Vol. 15 (1944), page 427.

[4] S. S WILKS, *Mathematical Statistics*, Princeton University Press, 1943

A SECOND FORMULA FOR THE PARTIAL SUM OF HYPERGEOMETRIC SERIES HAVING UNITY AS THE FOURTH ARGUMENT

By Hermann von Schelling

Naval Medical Research Laboratory, New London, Connecticut[1]

A convergent hypergeometric series with 1 as fourth argument has been expressed by Gauss, using gamma functions, as follows:

$$(1)\quad F(\alpha, \beta, \gamma; 1) = 1 + \frac{\alpha \cdot \beta}{\gamma \cdot 1} + \frac{\alpha(\alpha + 1)\ \beta(\beta + 1)}{\gamma(\gamma + 1) \cdot 1\ 2} + \cdots = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)}.$$

Let us denote the νth partial sum of $F(\alpha, \beta, \gamma; 1)$ by $F_\nu(\alpha, \beta, \gamma, 1)$, and let us put

$$(2)\qquad \frac{F_\nu(\alpha, \beta, \gamma; 1)}{F(\alpha, \beta, \gamma; 1)} = G_\nu(\alpha, \beta, \gamma).$$

The following equation is obvious·

$$(3)\qquad G_\nu(\alpha, \beta, \gamma) = G_\nu(\beta, \alpha, \gamma).$$

In [1] it is shown that

$$(4)\qquad G_\nu(\alpha, \beta, \gamma) = 1 - G_\alpha(\nu, \gamma - \beta - \alpha, \gamma - \alpha + \nu)$$

is valid if α is a positive integer.

If $(\gamma - \beta - \alpha)$ is a positive integer, (3) and (4) yield

$$\begin{aligned} G_\nu(\alpha, \beta, \gamma) &= 1 - G_\alpha(\gamma - \beta - \alpha, \nu, \gamma - \alpha + \nu) \\ &= G_{\gamma-\beta-\alpha}(\alpha, \beta, \alpha + \beta + \nu) \end{aligned}$$

In terms of partial sums of the hypergeometric series this becomes

$$(5)\qquad \begin{aligned} &\frac{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)}{\Gamma(\gamma)\Gamma(\gamma - \beta - \alpha)} F_\nu(\alpha, \beta, \gamma; 1) \\ &\qquad = \frac{\Gamma(\alpha + \nu)\Gamma(\beta + \nu)}{\Gamma(\nu)\Gamma(\alpha + \beta + \nu)} F_{\gamma-\beta-\alpha}(\alpha, \beta, \alpha + \beta + \nu; 1), \end{aligned}$$

which is a new formula involving partial sums of hypergeometric series with 1 as fourth argument. It is more useful than (4) if $\gamma - \beta - \alpha < \alpha$ or $\gamma < 2\alpha + \beta$.

It is of theoretic interest that the arguments of the new series do not depend on the third argument γ of the original series. Therefore it is possible to develop a simple recursion formula. If we write (5) for $(\gamma - 1)$ instead of γ, the series of the second member has one term less. Subtracting these equations yields after some simplifications

[1] Opinions or conclusions contained in this paper are those of the author. They are not to be construed as necessarily reflecting the views or endorsement of the Navy Department

$$
\begin{aligned}
&(\gamma - \alpha - 1)(\gamma - \beta - 1)F_\nu(\alpha, \beta, \gamma; 1) \\
(6) \qquad &\qquad - (\gamma - \beta - \alpha - 1)(\gamma - 1)F_\nu(\alpha, \beta, \gamma - 1; 1) \\
&\qquad\qquad = \frac{\Gamma(\nu + \alpha)}{\Gamma(\alpha)} \cdot \frac{\Gamma(\nu + \beta)}{\Gamma(\beta)} \cdot \frac{\Gamma(\gamma)}{\Gamma(\nu + \gamma - 1)} \cdot \frac{\Gamma(1)}{\Gamma(\nu)}.
\end{aligned}
$$

Many recursion formulas are known for hypergeometric functions, but (6) may be the first equation of this type linking two hypergeometric partial sums of ν terms each.

In order to demonstrate the numerical advantage of the new formula (5), we restate the example of [1]. An urn may contain N balls of which a black and b white A single ball is drawn. We note its color, return the ball into the urn and add Δ balls of the same color. The probability that the n_1th black ball appears at the latest in the n-th drawing is

$$
(7) \quad W(n) = \frac{\frac{a}{\Delta}\left(\frac{a}{\Delta} + 1\right) \cdots \left[\frac{a}{\Delta} + n_1 - 1)\right]}{\frac{N}{\Delta}\left(\frac{N}{\Delta} + 1\right) \cdots \left[\frac{N}{\Delta} + (n_1 - 1)\right]} F_{n-n_1+1}\left(n_1, \frac{b}{\Delta}, \frac{N}{\Delta} + n_1; 1\right).
$$

If $\frac{a}{\Delta}$ is a positive integer (5) yields

$$
\begin{aligned}
W(n) &= \frac{(n - n + 1)(n - n_1 + 2) \cdots n}{\left(\frac{b}{\Delta} + n - n_1 + 1\right)\left(\frac{b}{\Delta} + n - n_1 + 2\right) \cdots \left(\frac{b}{\Delta} + n\right)} \\
(8) \qquad & \qquad \cdot F_{a/\Delta}\left(n_1 \frac{b}{\Delta}, \frac{b}{\Delta} + n + 1; 1\right).
\end{aligned}
$$

If we take

$$
\Delta = 1, \qquad a = 1, \qquad b = N - 1,
$$

we get

$$
(9) \qquad W(n) = \frac{n!(N + n - n_1 - 1)!}{(n - n_1)!(N + n - 1)!}.
$$

Calculating $W(n)$, using the original formula (7), is quite tedious, but (5) sometimes simplifies the numerical work. Let us calculate the probability $W(6)$ that the third black ball appears in the 6th drawing, if the number of the original balls is $N = 10$. Using formulas (7), (4), and (9) respectively we have

$$
W(6) = \frac{3!9!}{12!}\left[1 + \frac{3 \cdot 9}{13 \cdot 1} + \frac{(3 \cdot 4)(9 \cdot 10)}{(13\ 14)(1 \cdot 2)} + \frac{(3 \cdot 4 \cdot 5)(9 \cdot 10 \cdot 11)}{(13 \cdot 14 \cdot 15)(1 \cdot 2\ 3)}\right] = \frac{4}{91},
$$

$$
W(6) = 1 - \frac{12!9!}{8!13!}\left[1 + \frac{4 \cdot 1}{14\ 1} + \frac{(4 \cdot 5)(1 \cdot 2)}{(14 \cdot 15)(1 \cdot 2)}\right] = \frac{4}{91};
$$

$$
W(6) = \frac{6!12!}{3!15!} = \frac{4}{91}.
$$

The time saved in using both formulas, of course, increases as the number of terms, $n - n_1 - 1$, of the original series, increases.

Let us mention that the special distribution corresponding to (9) does not have finite moments. For arbitrary values of N, a, Δ the arithmetic mean is

$$E(n) = \frac{N - \Delta}{a - \Delta} \cdot n_1, \tag{10}$$

the expectation of $n(n + 1)$ is

$$E[n(n + 1)] = \frac{(N - \Delta)(N - 2\Delta)}{(a - \Delta)(a - 2\Delta)} \cdot n_1(n_1 + 1), \tag{11}$$

and finally the variance

$$\sigma^2(n) = \frac{(N - \Delta)(N - a)[(n_1 - 1)\Delta + a]}{(a - \Delta)^2(a - 2\Delta)} \cdot n_1. \tag{12}$$

The mode can be derived from the fact that

$$w(n + 1) = w(n) \quad \text{for} \quad n = \frac{N}{a + \Delta} \cdot (n_1 - 1). \tag{13}$$

Especially we get $w(11) = w(10)$ for our numerical example.

The mean and variance do not exist for $a = \Delta = 1$, as in our example. However, it is possible to find a number n so that $W(n)$ takes any value near to unity, for instance .99. For large n and small n_1 (9) yields the approximation

$$W(n) = \frac{n(n - 1) \cdots (n - n_1 + 1)}{(N + n - 1)(N + n - 2) \cdots (N + n - n_1)}$$

$$\sim \left(\frac{n - \dfrac{n_1 - 1}{2}}{N + n - \dfrac{n_1 + 1}{2}} \right)^{n_1}.$$

Hence, $W(2666) = .99$ for our example. One needs 2666 trials if one wants a 99% probability for getting three black balls. This surprising result cannot be derived from the original formula (7).

REFERENCE

[1] H. VON SCHELLING, "A formula for the partial sums of some hypergeometric series", *Annals of Math. Stat.*, Vol. 20 (1949), pp. 120–122.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Chicago meeting of the Institute, April 28-29, 1950)

1. The Distribution of the Quotient of Ranges in Samples from a Rectangular Population. Paul R. Rider, Washington University, St. Louis, Missouri.

The distribution of the quotient of the ranges of two independent, random samples from a continuous rectangular population is derived The distribution is independent of the population range and can be used to test the hypothesis that two samples came from the same rectangular population just as the distribution of the variance ratio is used to test whether two samples came from the same normal population.

2. A Geometric Method for Finding the Distribution of Standard Deviations when the Sampled Population Is Arbitrary. (Preliminary Report). Paul Irick, Purdue University.

For an ordered random sample, $x_1 \leqq x_2 \leqq \cdots \leqq x_n$, chosen from a population, $f(x)$, $a \leqq x \leqq b$, let $r_i = x_{i+1} - x_i \geqq 0, i = 1, 2, \cdots, n - 1$. Make the transformation

$$r_i = -\sqrt{\frac{i-1}{2i}}\, r'_{i-1} + \sqrt{\frac{i+1}{2i}}\, r'_i,$$

and call U' the $1/n!$ portion of the r' space bounded by the $n - 1$ sphere and hyperplanes, $\sum_1^{n-1} r_i'^2 = 2ns^2$, $r'_i = \sqrt{\frac{i-1}{i+1}}\, r'_{i-1}$, $i = 1, 2, \cdots, n - 1$, where s is the sample standard deviation The point density in U', $\delta(r')$, is the transform of

$$\delta(r) = \int_{x_1=a}^{b-\Sigma r_i} f(x_1)f(x_1 + r_1) \cdots f(x_1 + r_1 + \cdots + r_{n-1})\, dx_1 .$$

Change to generalized polar coordinates and call U the outer hyperspherical boundary of U' whereon the density is designated by $\delta(\sqrt{2n}s, \varphi)$. Then $p(s)$, the probability law for s, is given by

$$p(s)\, ds = n!\, n^{n/2} s^{n-2}\, ds \int_{\varphi_1} \cdots \int_{\varphi_{n-2}} \delta(\sqrt{2n}s, \varphi) \sin^{n-3} \varphi_1 \cdots \sin \varphi_{n-3}\, d\varphi_{n-2} \cdots d\varphi_1,$$

where

$$\text{arc cos} \sqrt{\frac{n}{(n-i)(i+1)}} \leqq \varphi_i \leqq \text{arc cos} \left[\sqrt{\frac{i-1}{i+1}} \Big/ \tan \varphi_{i-1}\right], i = 1, 2, \cdots, n - 2,$$

whenever b is infinite. The distribution of sample range is readily found in U' and is expressible in the same form as $p(s)$ with the same limits of integration When b is finite, the complete integral holds only for $0 \leqq s \leqq (b - a)/\sqrt{2n}$, there being $n^2/4$ connected arcs in $p(s)$ if n is even, and $(n^2 - 1)/4$ arcs if n is odd. The axes are rotated to give relatively simple formulas for $p(s)$ when $n \leqq 4$, the case of $n = 5$ also being discussed. The method readily produces previously reported results for $p(s)$. In the application of the method, particular attention has been paid to the Type III and polynomial Type I populations The density function provides much information concerning the form of $p(s)$ for various populations, and contours of constant δ in U' are of theoretical interest.

3. Probability of a Correct Result with a Certain Rounding-off Procedure. W. S. LOUD, University of Minnesota

Consider the problem of the addition of n numbers expressed in the base B of numeration. Supposing each number known to arbitrary accuracy, to obtain the sum accurate to k places, one may round off each number to $(k + 1)$ places, add, and round the sum to k places If the numbers are assumed uniformly distributed, the probability that the above procedure gives the correct result may be found explicitly by use of characteristic functions. If the base B is odd, the result is $2(\pi B)^{-1} \int_0^\infty \sin^{n-1} u \sin^2 Bu \, u^{-n-1} \, du$, and if the base B is even, $2(\pi B)^{-1} \int_0^\infty \sin^2 Bu \cos u \, u^{-n-1} \, du$ Both formulas have the asymptotic formula $6^{1/2} B (\pi n)^{-1/2}$ as n becomes infinite.

4. Analysis of a One-person Game. (Preliminary Report). W. M. KINCAID, University of Michigan.

The problem of allocation of supplies is one which arises in many military and economic connections. The present report discusses a game constructed as a model of a simple situation of this type. The player is given a supply of cards, and receives payments for giving these up when certain random events occur during the period of play.

The optimal strategy, which maximizes the expected value of these payments, is governed by certain critical times such that the player's response to a particular event depends on whether it occurs before or after one of these times.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Dr. Leo A. Aroian, on leave from Hunter College, is acting as a Research Physicist in charge of computations at the Hughes Aircraft Co., Department of Electronics and Guided Missiles, Culver City, California.

Dr Ralph A. Bradley from McGill University, Montreal, Canada will join the staff as Associate Professor in the Department of Statistics at Virginia Polytechnic Institute on July 1, 1950. He will devote the majority of his time to research on rank order statistics.

Dr E. R. Dalziel has relinquished his post as Assistant Master at Technical School, New Zealand, to become Senior Engineer with the Overseas Telecommunication Commission, Australia.

On September 1, Dr. David Duncan from the University of Sydney, Sydney, Australia, will join the statistical staff of Virginia Polytechnic Institute as Associate Professor of Statistics. He will devote the majority of his time to teaching.

Dr. C. H. Fischer has been promoted to the rank of Professor of Actuarial Mathematics in the Department of Mathematics and Professor of Insurance in the School of Business Administration, University of Michigan, Ann Arbor, Michigan.

Dr. E. J Gumbel, Professor of Statistics at the New York New School for Social Research, has been appointed Consultant to the National Bureau of Standards and has been awarded a Guggenheim fellowship for finishing a book on the theory of extreme values.

Dr. Eugene Lukacs, who has been on leave from Our Lady of Cincinnati College and working as a Statistician for the U. S. Naval Ordnance Test Station, Inyokern, California, is transferring to the Statistical Engineering Laboratory, National Bureau of Standards, Washington, D. C

Dr. R. B. Leipnik, formerly a member of the Institute for Advanced Study, has accepted a position as Assistant Professor of Mathematics at the University of Washington, Seattle.

Mr. Harold C Mathisen, Jr., of the Kaiser-Frazer Corporation has been transferred from Willow Run, Michigan where he was an Assistant to the Director of Sales, to Buffalo, New York, as Regional Credit-Distribution Supervisor

Mr. Jack Moshman has resigned from the U. S. Atomic Energy Commission at Oak Ridge, Tennessee, to accept a position as Statistician with the Mathematics Panel of the Oak Ridge National Laboratory

Dr. D N. Nanda is now acting as Senior Scientific Officer in statistics at the Technical Development Estt. Laboratory at Kanpur, India.

Mr. Shanti A. Vora was awarded at the commencement June 5, 1950, the degree of Doctor of Philosophy in Mathematical Statistics from the University of North Carolina, Chapel Hill. His dissertation, entitled "Bounds on the Distribution of Chi-Square," won the William Chambers Coker Award in Science for 1950 granted by the Elisha Mitchell Scientific Society for excellence in research in all the scientific departments of the university. He has been appointed Acting Assistant Professor in the Department of Statistics at Stanford University, California, effective July 1, 1950, where he will be principally employed in research on sampling inspection

Professor Abraham Wald, Chairman, Department of Mathematical Statistics, Columbia University, gave a series of lectures on the theory of statistical decision functions at the Naval Ordnance Test Station, Inyokern, California, April 3–7, 1950. Representatives from several organizations and educational institutions on the Pacific coast attended the lectures.

A copy of the bulletin of the Graduate School of Public Health, University of Pittsburgh, has been received at the Secretary's office. The program of the Department of Biostatistics will be of particular interest to readers of the *Annals*. The teaching and research activities of the Department of Biostatistics are aimed primarily at the development of methods for the statistical appraisal of the health problems of groups: the community, the family, and the special aggregates such as the population in industry and in school.

The Educational Testing Service is offering for 1951–52 its fourth series of research fellowships in psychometrics leading to the Ph.D degree at Princeton University. Open to men who are acceptable to the Graduate School of the University, the two fellowships each carry a stipend of $2,375 a year and are normally renewable. Fellows will be engaged in part-time research in the general area of psychological measurement at the offices of the Educational Testing Service and will, in addition, carry a normal program of studies in the Graduate School. Competence in mathematics and psychology is a prerequisite for obtaining these fellowships. Information and application blanks may be obtained from: Director of Psychometric Fellowship Program, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

Preliminary Actuarial Examinations

Prize Awards

The winners of the prize awards offered by the Society of Actuaries to the nine undergraduates ranking highest on the score of Part 2 of the 1950 Preliminary Actuarial Examinations are as follows:

First Prize of $200

Mattuck, Arthur P.	Swarthmore College

Additional Prizes of $100

Dempster, Arthur P.	University of Toronto
Haslam, M. Brent	University of Buffalo
Hudek, Paul R.	University of Minnesota
Jamieson, J. Rae	University of Toronto
Leff, Milton M.	University of Western Ontario
Milnor, John W.	Princeton University
Reynolds, William F.	College of the Holy Cross
Walter, John R.	University of Toronto

The Society of Actuaries has authorized a similar set of nine prizes for the 1951 examinations on Part 2.

The Preliminary Actuarial Examinations consist of the following three examinations:

Part 1 *Language Aptitude Examination*
(Reading comprehension, meaning of words and word relationships, antonyms, and verbal reasoning)

Part 2. *General Mathematics Examination.*
(Algebra, trigonometry, coordinate geometry, differential and integral calculus.)

Part 3. *Special Mathematics Examination.*
(Finite differences, probability and statistics.)

The 1951 Preliminary Actuarial Examinations will be prepared by the Educational Testing Service and will be administered by the Society of Actuaries at centers throughout the United States and Canada on May 18, 1951. The closing date for applications is March 15, 1951.

Detailed information concerning the Examinations can be obtained from:

The Society of Actuaries
208 South LaSalle Street
Chicago 4, Illinois

New Members

The following persons have been elected to membership in the Institute

(March 1, 1950 to May 31, 1950)

Ard, Everett E., B.S (Kansas State Teachers College), Student, University of Michigan, *1657 Monson Court, Willow Run, Michigan*

Bainbridge, T. R., B S (Clemson College, S C), Supervisor, Koda Quality Inspection Group, Tennessee Eastman Corporation, Kingsport, Tennessee.

Bankier, James D., Ph D (Rice Institute), Associate Professor, Mathematics Department, McMaster University, Hamilton, Ontario, Canada.

den Broeder, Jr., George G., B S (Wayne Univ), Student, Wayne University, *459 East Grand Boulevard, Detroit 7, Michigan*

Casas, Luis T., Ph.D (Univ of Bogota, Colombia), Professor of Statistics, Universidad de los Andes and Facultad de Economia Industrial y Comercial del Gimnasio Moderno, also Statistician, Compania Colombiana de Seguros and Companie Colombiana de Seguros de Vida, *Apartado Nacional No. 2088, Bogota, Colombia.*

Clark, Charles R., B S. (Univ of Michigan), Student, University of Michigan, *1215 West Cross Street, Ypsilanti, Michigan*

Dolby, James L., M A (Wesleyan Univ.), Mathematical Physicist, Belding-Heminway, Inc., *66 Grove Street, Putnam, Connecticut*

Elfving, Gustav, Ph D (Helsingfors, Finland), Professor of Mathematics, University of Helsingfors, Finland, now visiting Professor, Mathematics Department, Cornell University, Ithaca, New York

Embody, Daniel R., M S (Cornell Univ), Staff Statistician, The Washington Water Power Company, *P O. Drawer 1445, Spokane 6, Washington.*

Frazier, David, Ph.D (Stanford Univ), Research Chemist, Chemical and Physical Research Division, The Standard Oil Company (Ohio), *2127 Cornell Road, Cleveland 6, Ohio.*

Graf, Herman S., B.A (Alfred Univ), Student, Department of Mathematical Statistics, University of North Carolina, *58 Winans Drive, Yonkers 2, New York*

Greenberg, Bernard G., Ph.D. (N C State College), Associate Professor and Acting Head, Department of Biostatistics, School of Public Health, Associate Professor, Institute of Statistics, Raleigh, North Carolina.

Grenander, Ulf, Ph D (Stockholm Univ.), Department of Mathematical Statistics, Norrtullsgatan 16, Stockholm, Sweden

Grosh, Jr., Louis E., M.S. (Purdue Univ), Research Assistant, Mathematics Department, Purdue University, W Lafayette, Indiana.

Hoffman, Walter, M A (Wayne Univ), Statistician, Research Laboratory, Childrens Fund of Michigan, *2903 Elmhurst, Detroit 6, Michigan.*

Hoffman, Robert G., A.B (Stanford Univ.), Student, University of Michigan, *420 Thompson Street, Ann Arbor, Michigan*

Hopkins, George D, B S. (Ohio State), Statistician, Sylvania Electric Products, Inc , Ottawa, Ohio, *404 N. Jameson Avenue, Lima, Ohio.*

Horowitz, Jacob, B S (Columbia Univ), Graduate Student, Department of Mathematical Statistics, Columbia University, *552 Riverside Drive, New York 27, New York*

Huntsberger, David V., M.S. (West Va Univ), Graduate Student, Iowa State College, Ames, Iowa, *224 Pammel Court, Ames, Iowa*

Kempff-Mercado, Rolando, Lic en Ciencias Econ (Univ Mayor de San Andres), Secretary General of Yacimientos Petroliferos Fiscales Bolivianos (Bolivian Oil Field Authority), *P O. Box 1283, La Paz, Bolivia*

Kennedy, Muriel E., B Sc. (University of Alberta), Statistician, Special Surveys Division, Dominion Bureau of Statistics, *128 Mason Terrace, Ottawa, Ontario, Canada*

Lander, Elmer L., B A (Western Reserve Univ , Cleveland), Student, University of Michigan, [illegible] *Cleveland 20, Ohio*

Li-Min, Tang, M.S. (Univ. of Mich.), Student, University of Michigan, *1109 Willard, Ann Arbor, Michigan*

Lin, Shao-kung, M A (Louisiana State Univ.), Student, Department of Economics, University of Illinois, *1202½ W. University Avenue, Urbana, Illinois.*

Mandelson, Joseph, B.S. (College of City of New York), Mathematical Statistician, Chief, Quality Assurance Branch, Inspection Division, Office of Chief, Army Chemical Center, Maryland, *80 Cedar Street, Edgewood Heights, Maryland.*

Marthens, Arthur S., B S. (Carnegie Inst. of Tech.), Mathematical Statistician, Bureau of Ships, Navy Department, Washington, D. C., *1820 Montier Street, Wilkinsburg, Pennsylvania.*

Masel, Marvin, A M (Columbia Univ.), Engineering Statistician, Goodyear Aircraft Corp., Akron, Ohio, *Y.M C A , Room 809, 80 Center Street, Akron, Ohio.*

McCune, Duncan C., B A (College of Wooster), Graduate Assistant, Mathematics Department, Purdue University, Lafayette, Indiana

Meyer, Paul L., B S. (Univ of Washington), Research Fellow, Laboratory of Mathematical Statistics, University of Washington, *9200-5th, N.E., Seattle 5, Washington.*

Milberg, Stanley, M A. (Columbia Univ), Statistician, *220 Turrell Avenue, So Orange, New Jersey.*

Miser, Hugh J., Ph.D. (Ohio State Univ), Operations Analyst Headquarters, United States Air Force, *2713 Blaine Drive, Chevy Chase 15, Maryland*

Morrison, Milton, M.A (Columbia Univ.), Instructor of Mathematics, Stevens Institute of Technology, Hoboken, New Jersey

Mulholland, Hugh P., Ph.D (Cambridge Univ , England), Associate Professor of Mathematics, American University of Beirut, Beirut, Lebanon.

Nelson, A. Carl, M.S. (Univ. of Delaware), Instructor in Mathematics, University of Delaware, Newark, *Marshallton, Delaware.*

Neuwirth, Sidney I., B.A (N Y. Univ.), Statistician, Biological Research Laboratories, Schering Corporation, *86 Orange Street, Bloomfield, New Jersey*

Pierce, James A., B.A. (Westminster College, Fulton, Mo), Graduate Assistant, Purdue University, *205 Sylvia, West Lafayette, Indiana.*

Powell, Claude J., B.S. (Univ of Tennessee), Quality Control Engineer, North American Rayon Corp , *615 Hattie Avenue, Elizabethton, Tennessee.*

Rojas, Basilio A., B S. (National College of Agriculture, Mexico), Graduate Student, Iowa State College, Statistical Laboratory, Ames, Iowa

Roseboom, John H., M.S. (Dartmouth College), Instructor, Department of Economics, Indiana University, Bloomington, Indiana.

Sandlin, William T., A B (Marshall College, Huntington, W Va), Independent Sales Engineer, *20 Fairfax Drive, Huntington, West Virginia*

Schmitt, Samuel A., B.S. (Univ. of Chicago), Research Analyst, Department of Defense, Washington, D. C., *1410 N Rhodes Street, Arlington, Virginia.*

Shaw, Richard H., M.S (Purdue Univ.), Research Fellow, Purdue University, *F.P H.A. 513-1 Airport Road, West Lafayette, Indiana.*

Smith, Hugh F., M.S.A (Cornell Univ.), Professor of Experimental Statistics, Institute of Statistics, University of North Carolina, *Box 5457, College Station, Raleigh, North Carolina.*

Sommers, Lysle D., B.S. (Bowling Green State Univ.), Sampling Assistant, Survey Research Center, University of Michigan and Graduate Student, *1532 Leeds Court, Willow Run, Michigan*

Springer, Clifford H., M Sc (Purdue Univ), Instructor, Department of Mathematics, Research Assistant, Statistical Laboratory, Recitation Building, Purdue University, West Lafayette, Indiana

Stearman, Robert L., M S (Oregon State College), Teaching Fellow, Department of Mathematics, Oregon State College, Corvallis, Oregon.

Tingey, Fred H., M S (Univ of Washington), Research Associate, University of Washington, *3310 Goldendale Place, Seattle 5, Washington*

Tipton, Lamar B., M A. (Columbia Univ), Statistical Clerk, Standard Oil Co of New Jersey, *6 W. 604th Street, Shanks Village, Orangeburg, New York*

Topp, Chester W., M A (Univ of Illinois), Associate Professor of Mathematics, Fenn College, *1524 Comton Road, Cleveland Heights 18, Ohio.*

Vora, Shanti A., M Sc (Bombay), Student, Department of Mathematical Statistics, University of North Carolina, *210 A, Phillips Hall, Chapel Hill, North Carolina*

Willis, Myron J., A.M (Indiana Univ), Instructor of Mathematics, Purdue University, Statistical Laboratory, Lafayette, Indiana.

REPORT OF THE CHICAGO MEETING OF THE INSTITUTE

The forty-third meeting and the first regional Mid-western meeting of the Institute of Mathematical Statistics was held on the campus of the University of Chicago, Chicago, Illinois on Friday and Saturday, April 28 and 29, 1950. The morning session on April 29 was held jointly with the American Mathematical Society. The following forty-six members of the Institute were registered as present:

K J. Arnold, Max Astrachan, Reinhold Baer, Alvin G Brooks, I W Burr, P G. Carlson, Herman Chernoff, P S Dwyer, H P Evans, J S Frame, Mary Goins, R D Gordon, John Gurland, P R Halmos, P C. Hammer, W L Hart, M A Hatke, P. E Irick, Howard L. Jones, Leo Katz, J P. Kelly, W M Kincaid, L. A Knowler, Tjalling Koopmans, F C. Leone, F W. Lott, W G. Madow, A. M. Mark, John W. Mauchly, Kenneth May, Duncan C. McCune, Cyril G Peckham, G B Price, P. R Rider, Norman Rudy, L J Savage, G. R Seth, Richard H. Shaw, Jack Sherman, M D Springer, Robert G D Steel, Z Szatrowski, J V Talacko, R. M. Thrall, L M Weiner, M E Wescott.

Professor Lloyd A. Knowler of the University of Iowa presided at the Friday afternoon session. The program consisted of the following invited papers:

1. *Why and Where Should Courses in Statistics Be Offered to Engineering Students?* M. E. Wescott, Northwestern University
2. *What and How Statistics Should be Taught to Engineering Students* I W. Burr, Purdue University.

Following this session a tea was given by the Department of Mathematics of the University of Chicago

Professor John Gurland of the University of Chicago presided at the Saturday morning session. This session was held jointly with the American Mathematical Society. The program was as follows:

1. *The Distribution of the Quotient of Ranges in Samples From a Rectangular Population* Paul R Rider, Washington University, St. Louis, Missouri
2. *A Geometric Method for Finding the Distribution of Standard Deviations when the Sampled Population Is Arbitrary* (Preliminary report) Paul Irick, Purdue University.
3. *Probability of a Correct Result with a Certain Rounding-off Procedure* W S Loud, University of Minnesota
4. *Analysis of a One-person Game.* (Preliminary report) W. M Kincaid, University of Michigan.

Professor W. G. Madow of the University of Illinois presided at the Saturday afternoon session. The program consisted of the following invited papers:

1 *Correlation and Regression with Matrix Factorization*. P. S. Dwyer, University of Michigan.
2. *The Identification of Structural Characteristics*. Tjalling Koopmans, University of Chicago, and Olav Reiersøl, University of Oslo, Norway.

K. J. Arnold,
Associate Secretary

THE PROBLEM OF THE GREATER MEAN

By Raghu Raj Bahadur and Herbert Robbins[1]

University of North Carolina

1. Introduction and summary. Let π_1, π_2 be normal populations with means m_1, m_2 respectively and a common variance σ^2, the parameter point $\omega = (m_1, m_2 : \sigma)$ which characterizes the two populations being unknown, and let Ω be an arbitrary given set of possible points ω. Random samples of fixed sizes n_1, n_2 are drawn from π_1, π_2 respectively, giving the combined sample point $v = (x_{11}, x_{12}, \cdots, x_{1n_1}, x_{21}, x_{22}, \cdots, x_{2n_2})$. For reasons which will be made clear later in connection with practical examples, any function $f(v)$ such that $0 \leq f(v) \leq 1$ is called a *decision function*, and for any such $f(v)$ the *risk function* is defined to be

$$r(f \mid \omega) = \max [m_1, m_2] - m_1 E[f \mid \omega] - m_2 E[1 - f \mid \omega] \geq 0, \tag{1}$$

where E denotes the expectation operator A decision function $\bar{f}(v)$ is said to be (a) *uniformly better* than $f(v)$ if $r(\bar{f} \mid \omega) \leq r(f \mid \omega)$ for all ω in Ω, the strict inequality holding for at least one ω, (b) *admissible* if no decision function is uniformly better than $\bar{f}(v)$, and (c) *minimax* if

$$\sup_{\omega \in \Omega} [r(\bar{f} \mid \omega)] = \inf_f \sup_{\omega \in \Omega} [r(f \mid \omega)].$$

The "problem of the greater mean" is, for any given Ω, to determine the minimax decision functions, particularly those which are also admissible. Special interest attaches to the case in which there exists a *unique* minimax decision function $\bar{f}(v)$ (in the sense that if $f(v)$ is any minimax decision function then $f(v) = \bar{f}(v)$ for almost every v in the sample space); such an $\bar{f}(v)$ is automatically admissible.

The problem of the greater mean is, of course, a special problem in Wald's general theory of statistical decision functions [1]. Our results will, however, be derived by very simple direct methods which make no use of Wald's general theorems.

We cite without proofs a few examples in order to show how strongly the solution of the problem of the greater mean depends on the structure of Ω. In each case the minimax decision function is a function only of the two sample means $\bar{x}_1$, $\bar{x}_2$.

(i) Let Ω' consist of the two points $(a, b: \sigma)$ and $(b, a: \sigma)$, with $a < b$ Then

$$f^*(v) = \begin{cases} 1 \text{ if } n_1\bar{x}_1 - n_2\bar{x}_2 > (n_1 - n_2)(a + b)/2, \\ 0 \text{ otherwise,} \end{cases} \tag{2}$$

is the unique minimax decision function.

[1] This work was supported in part by the Office of Naval Research.

(ii) Let Ω'' consist of the two points $(c + h, c: \sigma)$ and $(c - h, c: \sigma)$, with $h > 0$ Then

$$f_c^0(v) = \begin{cases} 1 \text{ if } \bar{x}_1 > c, \\ 0 \text{ otherwise,} \end{cases} \tag{3}$$

is the unique minimax decision function.

(iii) Let Ω''' consist of the three points $(\frac{1}{2}, -\frac{1}{2}:1)$, $(\frac{1}{2}, \frac{3}{2}:1)$, $(-\frac{3}{2}, -\frac{1}{2}:1)$, and let $n_1 = n_2 = n$. Then

$$f^{**}(v) = \begin{cases} 1 \text{ if } e^{-2n\bar{x}_1} + e^{2n\bar{x}_2} < \lambda, \\ 0 \text{ otherwise,} \end{cases} \tag{4}$$

where λ is a certain definite constant, is the unique minimax decision function.

The parameter spaces of two or three points specified in these examples are rather trivial, but in fact the corresponding decision functions (2), (3), (4) remain the unique minimax solutions of the decision problem with respect to much more general parameter spaces. Thus, for example, it is clear that $f^*(v)$ will remain the unique minimax decision function with respect to any Ω which contains Ω' and is such that

$$\sup_{\omega \in \Omega} [r(f^* \mid \omega)] = \sup_{\omega \in \Omega'} [r(f^* \mid \omega)].$$

Corresponding remarks apply to $f_c^0(v)$ and $f^{**}(v)$.

When $n_1 = n_2$, (2) reduces to

$$f^0(v) = \begin{cases} 1 \text{ if } \bar{x}_1 > \bar{x}_2, \\ 0 \text{ otherwise.} \end{cases} \tag{5}$$

This decision function is of particular interest when both the means m_1, m_2 are unknown. It will be shown that whether or not $n_1 = n_2$, $f^0(v)$ is the unique minimax decision function under certain conditions on Ω which are likely to hold in practice, at least when both n_1 and n_2 are sufficiently large (Theorem 3) Likewise, $f_c^0(v)$, which is the analogue of $f^0(v)$ when one of the means (m_2) is known exactly, is apt to be the unique minimax decision function in such cases, at least when n_1 is sufficiently large (Theorem 4). These results on $f^0(v)$ and $f_c^0(v)$ form the main results of the present paper.

So much by way of a general summary. We shall now give a practical illustration (another is given in Section 3) to show how the problem of the greater mean arises in applications.

Suppose that a consumer requires a certain number of manufactured articles which can be supplied at the same cost by each of two sources π_1 and π_2. The quality of an article is measured by a numerical characteristic x, and it is known that in the product of π_i, x is normally distributed with mean m_i and variance σ^2, but the values of these parameters are unknown. The consumer has obtained a random sample of n_1 and n_2 articles from π_1 and π_2 respectively, and has found the values of x to be $(x_{11}, x_{12}, \cdots, x_{1n_1}, x_{21}, x_{22}, \cdots, x_{2n_2}) = v$. What is the best way of ordering a total of N articles from the two sources?

The usual statistical theory, which confines itself to estimating the unknown parameters and to testing hypotheses of the form $H_0(m_1 = m_2)$, has at best an indirect bearing on the problem at hand. We therefore adopt Wald's point of view and investigate the consequences of any given course of action. If the consumer orders fN articles from π_1 and $(1 - f)N$ from π_2, where $0 \leq f \leq 1$, then the expectation of the sum of the x-values in the articles he obtains will be $N(m_1 f + m_2(1 - f))$ The maximum possible value of this quantity is $N \max [m_1, m_2]$, and the "loss" per article which he sustains may therefore be taken as

$$W(\omega, f) = \max [m_1, m_2] - m_1 f - m_2(1 - f) \geq 0,$$

where $\omega = (m_1, m_2 : \sigma)$ is the true parameter point.

The consumer wants to choose f so as to make W as small as possible. If he knew m_1 to be greater, or to be less, than m_2, then by choosing $f = 1$ or 0 respectively he could make $W = 0$ But since he does not know which m_i is the greater he will presumably choose f as some function of the sample point v. Suppose, therefore, that a "decision function" $f(v)$, such that $0 \leq f(v) \leq 1$ but not necessarily taking on only the values 0 and 1, is defined for all points v in the sample space and that the consumer sets $f = f(v)$.[2] In repeated applications of this procedure, the "risk" or expected loss (a double expectation is involved: the expected loss for a given f and the expected value of f in using the decision function $f(v)$) per article is given by (1), and the consumer will try to find an $f(v)$ which minimizes this risk. Since the value of the risk depends on ω it is necessary to specify which values of ω are to be regarded as possible in the given problem; let the set of all such ω be denoted by Ω. If the consumer agrees to adopt the "conservative" criterion of minimizing the maximum possible risk, then the statistician's problem is to find the minimax decision functions in the sense defined above We have given the solutions of this problem for certain types of parameter spaces. The reader will observe that each of the minimax decision functions (2), (3), (4) was of the "all or nothing" type, with values 0 and 1 only. (Whether this remains true for every Ω we do not know.) By using one of these decision functions in a given instance one arrives at either the best possible decision or the worst. The attitudes of doubt sometimes associated with the non-rejection of the hypothesis $H_0(m_1 = m_2)$ are therefore

---

[2] One might say that the consumer should choose f in the light of what he can infer from v about the m_i. But this formulation as a problem in ordinary statistical inference (estimation and testing) is not relevant and may be misleading For example, a plausible $f(v)$, based on the idea that the problem is one of testing hypotheses, is as follows· "Perform the two-tailed t test of $H_0(m_1 = m_2)$ at the five per cent level If H_0 is rejected set $f = 0$ or 1 according as $\bar{x}_1$ is less than or greater than $\bar{x}_2$. If H_0 is not rejected set $f = \frac{1}{2}$." Another $f(v)$, based on the theory of estimation, according to which the $\bar{x}_i$ are the "best" estimates of the m_i, is as follows "Set $f = 0$ or 1 according as $\bar{x}_1$ is less than or greater than $\bar{x}_2$." Actually, the latter procedure is, from the remarks above concerning (5), the "best" in a certain definite sense and under certain conditions, but this fact does not follow from the usual theory of estimation

irrelevant to the problem of the greater mean in the examples cited. (Cf. footnote 2; also Example 1 in Section 3.)

The risk function (1) is but one of a general class R of risk functions, to be defined in Section 2, which are associated with the problem of the greater mean. The most important members of R are (1) and

$$\bar{r}(f \mid \omega) = P(\text{incorrect decision using } f(v) \mid \omega), \tag{6}$$

where "$m_1 \leq m_2$" and "$m_1 \geq m_2$" are the two possible decisions. The risk function (6) is relevant to applications of a purely "scientific" nature in which the statistician is asked merely to give his opinion as to which population has the greater mean. Although the problem of constructing a suitable decision function for (6) is akin in spirit to the problems considered in the now classical Neyman-Pearson theory of statistical tests, no satisfactory solutions seem to be available. It is easy to see, however, that (1) and (6) are quite similar. Of course, in the case of (1) a decision function $f(v)$ may take on any value between 0 and 1 inclusive, while for (6) we allow only functions which take on only the values 0 and 1, corresponding respectively to the decisions "$m_1 \leq m_2$" and "$m_1 \geq m_2$". We then have for any such $f(v)$,

$$\bar{r}(f \mid \omega) = \begin{cases} P(f(v) = 1 \mid \omega) = E[f \mid \omega] & \text{if } m_1 < m_2, \\ P(f(v) = 0 \mid \omega) = E[1 - f \mid \omega] & \text{if } m_1 > m_2, \\ 0 & \text{if } m_1 = m_2 \end{cases} \tag{6'}$$

and by comparison with (1) we see that $r(f \mid \omega) = |m_1 - m_2| \bar{r}(f \mid \omega)$ for all ω. Now, in the three examples (i), (ii), (iii) cited above the unique minimax decision functions happen to take on only the values 0 and 1, and $|m_1 - m_2|$ is constant on each of the respective parameter sets. It follows that (2), (3), (4) are also the unique minimax decision functions relative to (6) and to Ω', Ω'', Ω''' respectively. The remarks above following Example (iii) also remain valid for the risk function (6).

We conclude this section with a remark on the methods of this paper Any decision function relevant to (6) is equivalent to a test of the hypothesis $H_0(m_1 < m_2)$ against the alternative $H_1(m_1 > m_2)$, the region $\{v : f(v) = 1\}$ being the "critical region." Hence the Neyman-Pearson probability ratio method can be used to obtain the unique minimax decision function with respect to (6) and an Ω consisting of two (or more) points, and the result carries over to more general types of Ω in the manner already indicated. It turns out, however, that the dominant properties of the probability ratio tests are not confined to the class of tests alone, but extend to the class of all functions $f(v)$ such that $0 \leq f(v) \leq 1$. This result (Theorem 1) enables us to solve the problem of the greater mean for the risk function (1) as well as for (6). The reader who is interested in applications may turn to Section 3.

2. Theorems. We require the following slight generalization of a well-known result of Neyman and Pearson [2]

THEOREM 1. *Let $\phi(v)$, $\phi_1(v)$, $\phi_2(v)$, $\cdots$, $\phi_r(v)$ be summable functions defined on a measure space E with points v and measure μ, $\mu(E) \leq \infty$, let $c_1, \cdots, c_r$ be arbitrary constants, and let $A \subseteq E$ be such that*

$$\text{(7)} \qquad \begin{cases} v \in A \text{ implies } \phi(v) \geq \sum_1^r c_i\phi_i(v), \\ v \in E - A \text{ implies } \phi(v) \leq \sum_1^r c_i\phi_i(v). \end{cases}$$

Set

$$\text{(8)} \qquad \int_A \phi_i \, d\mu = a_i \qquad (i = 1, \cdots, r),$$

and let $f(v)$ be any measurable function such that

$$\text{(9)} \qquad 0 \leq f(v) \leq 1$$

and such that

$$\text{(10)} \qquad \int_E f\phi_i \, d\mu = a_i \qquad (i = 1, \cdots, r).$$

Then

$$\text{(11)} \qquad \int_E f\phi \, d\mu \leq \int_A \phi \, d\mu.$$

PROOF.

$$\int_E f\phi \, d\mu = \int_A f\phi \, d\mu + \int_{E-A} f\phi \, d\mu$$

$$\leq \int_A f\phi \, d\mu + \int_{E-A} f\left(\sum_1^r c_i\phi_i\right) d\mu \qquad \text{by (9), (7),}$$

$$= \int_A f\phi \, d\mu + \sum_1^r c_i \int_{E-A} f\phi_i \, d\mu$$

$$= \int_A f\phi \, d\mu + \sum_1^r c_i \left[\int_E f\phi_i \, d\mu - \int_A f\phi_i \, d\mu\right]$$

$$= \int_A f\phi \, d\mu + \sum_1^r c_i \left[a_i - \int_A f\phi_i \, d\mu\right] \qquad \text{by (10),}$$

$$= \int_A f\phi \, d\mu + \sum_1^r c_i \left[\int_A (1 - f)\phi_i \, d\mu\right] \qquad \text{by (8),}$$

$$= \int_A \phi \, d\mu - \int_A (1 - f)\phi \, d\mu + \int_A (1 - f) \left(\sum_1^r c_i\phi_i\right) d\mu$$

$$= \int_A \phi \, d\mu + \int_A (1 - f) \left(\sum_1^r c_i\phi_i - \phi\right) d\mu$$

$$\leq \int_A \phi \, d\mu \qquad \text{by (9), (7).}$$

NOTE 1. *If the condition*

$$\mu\left\{v : \phi(v) = \sum_1^r c_i \phi_i(v)\right\} = 0 \tag{12}$$

holds, then in order that the equality hold in (11) *it is necessary and sufficient that*

$$f(v) = \chi_A(v) \qquad a.e.\ (\mu), \tag{13}$$

where $\chi_A(v)$ is the characteristic function of the set A,

$$\chi_A(v) = \begin{cases} 1 & \text{if } v \in A, \\ 0 & \text{if } v \in E - A. \end{cases}$$

PROOF. The sufficiency is obvious. To prove the necessity we observe from the proof of Theorem 1 that for equality to hold in (11) it is necessary that

$$f(v)\left(\phi\ (v) - \sum_1^r c_i \phi_i(v)\right) = 0 \qquad \text{a.e. } (\mu) \text{ in } E - A,$$

and that

$$(1 - f(v))\left(\phi(v) - \sum_1^r c_i \phi_i(v)\right) = 0 \qquad \text{a.e. } (\mu) \text{ in } A.$$

These relations and (12) imply (13).

NOTE 2. *If relations* (10) *are replaced by*

$$\int_E f\phi_i\, d\mu \le a_i \qquad (i = 1, \cdots, r), \tag{10'}$$

and if each of the constants c_i is non-negative, then Theorem 1 *and Note* 1 *remain valid.*

Theorem 1 has applications to a number of decision problems of a certain type. In the present paper we consider only the "problem of the greater mean" for two normal populations with a common variance σ^2, where at least one of the means m_1, m_2 is unknown. The following assumptions and definitions will be valid henceforth.

(A) E_N is the $N = n_1 + n_2$ dimensional sample space of points $v = (x_{11}, x_{12}, \cdots, x_{1n_1};\ x_{21}, x_{22}, \cdots, x_{2n_2})$. A measurable function $f(v)$ defined for all v in E_N is a *decision function* if $0 \le f(v) \le 1$. $f_1(v) \equiv f_2(v)$ means $f_1(v) = f_2(v)$ for almost every v in E_N.

(B) Ω is a given set of points $\omega = (m_1, m_2 : \sigma)$, $\sigma > 0$. Given ω in Ω, the probability measure in E_N is that generated by the distribution function

$$K(v \mid \omega) = \prod_{i=1}^{2} \prod_{j=1}^{n_i} G\left[(x_{ij} - m_i)/\sigma\right],$$

where

$$G(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-u^2/2}\, du.$$

Given any function $\phi = \phi(v)$ for which the integral exists we write

$$E[\phi \mid \omega] = \int_{E_N} \phi(v)\, dK(v \mid \omega).$$

(C) Let $\gamma(\omega) = (g_1, g_2)$ be a function defined for all ω in Ω, with values in E_2, and such that

$$m_i \leq m_j \quad \text{implies} \quad g_i \leq g_j \qquad (i, j = 1, 2). \tag{14}$$

Given p, $0 \leq p \leq 1$, we define

$$W(\omega, p) = \max[g_1, g_2] - g_1 p - g_2(1 - p),$$

and given a decision function $f(v)$ we define the *risk function*

$$\begin{aligned} r(f \mid \omega) &= E[W(\omega, f) \mid \omega] = W(\omega, E[f \mid \omega]) \\ &= \max[g_1, g_2] - g_1 E[f \mid \omega] - g_2 E[1 - f \mid \omega] \end{aligned} \tag{15}$$

The class of risk functions (15) corresponding to all functions $\gamma(\omega)$ which satisfy (14) is denoted by R (The two most important members of R are (1), with

$$\gamma(\omega) = (m_1, m_2),$$

and (6), with

$$\gamma(\omega) = \begin{cases} (0, 1) & \text{if} \quad m_1 < m_2, \\ (1, 0) & \text{if} \quad m_1 > m_2, \\ (0, 0) & \text{if} \quad m_1 = m_2 \end{cases}$$

The risk functions (1) and (6) appear in the examples in Section 3.) Throughout this section $r(f \mid \omega)$ will denote a fixed but arbitrary member of R. We shall use the notations

$$\begin{aligned} h(\omega) &= | g_1 - g_2 |, \\ d(\omega) &= \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} (m_1 - m_2)/\sigma, \\ \bar{x}_i &= n_i^{-1} \sum_{j=1}^{n_i} x_{ij} \qquad (i = 1, 2). \end{aligned}$$

THEOREM 2. *Let $\omega_1 = (m_1, m_2 : \sigma)$ and $\omega_2 = (\mu_1, \mu_2 : \sigma)$ be two parameter points such that*

$$d(\omega_1) < 0, \qquad d(\omega_2) > 0, \qquad h(\omega_1)h(\omega_2) > 0.$$

For any λ, $-\infty \leq \lambda \leq \infty$, let $f_\lambda(v)$ be the characteristic function of the set

$$A_\lambda = \{v : n_1(\mu_1 - m_1)\bar{x}_1 + n_2(\mu_2 - m_2)\bar{x}_2 > \lambda\sigma\}. \tag{16}$$

Then

(i) *Corresponding to any decision function $f(v)$, there exists a λ such that*

$$r(f_\lambda \mid \omega_1) = r(f \mid \omega_1), \qquad r(f_\lambda \mid \omega_2) \leq r(f \mid \omega_2);$$

the inequality is strict unless $f(v) \equiv f_\lambda(v)$.

(ii) *Given any* λ, *if* $f(v)$ *is a decision function such that*

$$r(f \mid \omega_i) \leq r(f_\lambda \mid \omega_i) \qquad (i = 1, 2),$$

then

$$f(v) \equiv f_\lambda(v).$$

(iii) *There exists a unique* c *such that*

$$r(f_c \mid \omega_1) = r(f_c \mid \omega_2) = B \text{ say,} \tag{17}$$

and for any decision function $f(v)$ *we have*

$$B \leq \max\,[r(f \mid \omega_1), r(f \mid \omega_2)]; \tag{18}$$

the inequality is strict unless $f(v) \equiv f_c(v)$. *It follows that* $f_c(v)$ *is the unique minimax decision function corresponding to the two-point parameter space* $\Omega = (\omega_1, \omega_2)$

Proof[3] (a) Let $\phi(v)$, $\phi_1(v)$ be the joint frequency functions of the sample point v corresponding to the parameter points ω_2, ω_1 respectively. It is readily seen that for any λ there exists a unique constant $c_1(\lambda)$, $0 \leq c_1(\lambda) \leq \infty$, such that

$$A_\lambda = \{v : \phi(v) > c_1\phi_1(v)\}$$

$(c_1(-\infty) = 0, c_1(\infty) = \infty)$. Moreover, since $\omega_1 \neq \omega_2$,

$$\mu\{v : \phi(v) = c_1\phi_1(v)\} = 0.$$

It follows from Theorem 1, Note 2, that if $f(v)$ is any decision function such that

$$E[f \mid \omega_1] \leq E[f_\lambda \mid \omega_1],$$

then

$$E[f \mid \omega_2] \leq E[f_\lambda \mid \omega_2],$$

and the strict inequality holds unless $f(v) \equiv f_\lambda(v)$.

(b) It is clear from the definition (16) that for any fixed parameter point ω the function

$$E[f_\lambda \mid \omega] = P(A_\lambda \mid \omega)$$

is continuous and strictly decreasing from 1 to 0 as λ varies from $-\infty$ to $+\infty$.

(c) For any decision function $f(v)$ and any parameter point ω we have by (C),

$$r(f \mid \omega) = \max\,[g_1, g_2] - g_1E[f \mid \omega] - g_2E[1 - f \mid \omega].$$

Hence

$$\begin{cases} r(f \mid \omega_1) = h(\omega_1)E[f \mid \omega_1], & h(\omega_1) > 0, \\ r(f \mid \omega_2) = h(\omega_2)E[1 - f \mid \omega_2], & h(\omega_2) > 0. \end{cases} \tag{19}$$

---

[3] Theorem 2 (as also Example (iii) of Section 1) could be derived from Wald's general results on the completeness of the class of Bayes solutions of statistical decision problems.

Since for any decision function $f(v)$, $0 \leq E[f \mid \omega_1] \leq 1$, we can by (b) choose λ so that

$$E[f_\lambda \mid \omega_1] = E[f \mid \omega_1], \tag{20}$$

and by (a) it follows that unless $f(v) \equiv f_\lambda(v)$,

$$E[f_\lambda \mid \omega_2] > E[f \mid \omega_2] \tag{21}$$

(i). Follows from (19), (20) and (21)
(ii). Follows from (19) and (a).
(iii). (17) follows from (19) and (b). Then (18) follows from (17) and (ii).

Theorem 2 provides the solution of any problem of the greater mean when Ω consists of just two points ω_1, ω_2. For, the problem is trivial unless $d(\omega_1)\, d(\omega_2) < 0$ and $h(\omega_1)h(\omega_2) > 0$, and in the non-trivial case the unique minimax decision function is $f_c(v)$ defined by (17). Moreover, it follows at once from the definition that if $\bar{f}(v)$ is the unique minimax decision function with respect to some parameter set $\bar{\Omega}$, then it remains so with respect to any Ω such that $\Omega \supseteqq \bar{\Omega}$ and

$$\sup_{\omega \in \Omega} \quad [r(\bar{f} \mid \omega)] = \sup_{\omega \in \bar{\Omega}} \quad [r(\bar{f} \mid \omega)]$$

By taking sets $\bar{\Omega}$ which consist of two points, Theorem 2 can therefore be used to obtain sufficient conditions for an $\bar{f}(v) = f_c(v)$ to be the unique minimax decision function with respect to a quite general Ω. (It is clear that results analogous to Theorem 2(iii) but pertaining to more than two parameter points can be derived from Theorem 1, and that these results can be exploited in a similar way. An instance of this procedure where $\bar{\Omega}$ consists of three points will be given at the end of this section.)

The theorems which follow exploit Theorem 2 in this way to obtain conditions on Ω under which the decision functions $f^0(v)$ and $f_c^0(v)$ defined by (5) and (3) are minimax. We consider $f^0(v)$ first. From (C) we have, after a simple computation,

$$r(f^0 \mid \omega) = h(\omega) \cdot G(- \mid d(\omega) \mid). \tag{22}$$

THEOREM 3. *Suppose that there exist sequences* $\{\omega_k\}$, $\{\omega_k'\}$ *of points* $\omega_k = (m_{1k}, m_{2k} : \sigma_k)$, $\omega_k' = (\mu_{1k}, \mu_{2k} : \sigma_k)$ *in* Ω *such that*

(i) $\lim_{k \to \infty} r(f^0 \mid \omega_k) = \sup_{\omega \in \Omega} [r(f^0 \mid \omega)] \qquad (\neq 0, \infty)$,

(ii) $d(\omega_k) = - d(\omega_k')$, $h(\omega_k) = h(\omega_k')$, *and* $n_1 m_{1k} + n_2 m_{2k} = n_1 \mu_{1k} + n_2 \mu_{2k}$ *for every* $k = 1, 2, \cdots$.

Then $f^0(v)$ *is an admissible minimax decision function. If there exist* $\omega_0 = (m_1, m_2 : \sigma)$, $\omega_0' = (\mu_1, \mu_2 : \sigma)$ *in* Ω *satisfying* (i) *and* (ii), *then* $f^0(v)$ *is the unique minimax decision function.*

PROOF. By (22) and (ii),

$$r(f^0 \mid \omega_k) = r(f^0 \mid \omega_k') \text{ for every } k. \tag{23}$$

Without loss of generality, we may assume the two sequences to be so chosen that $h(\omega_k) = h(\omega_k') > 0$ for every k. Then, by interchanging corresponding members if necessary, we may assume that

$$d(\omega_k) = -d(\omega_k') < 0 \text{ for every } k. \tag{24}$$

Consider the two points ω_k, ω_k' in Ω with arbitrary but fixed k. Writing ω_k, ω_k' for ω_1, ω_2 respectively, and using conditions (ii), a simple calculation shows that the set defined by (16) is

$$A_\lambda = \{v : \bar{x}_1 - \bar{x}_2 > L\}, \tag{25}$$

L being a strictly increasing function of λ.

Choose and fix an arbitrary decision function $f(v) \not\equiv f^0(v)$. Comparing (5) and (25), it follows from Theorem 2(iii) and (23) that

$$r(f^0 \mid \omega_k) = r(f^0 \mid \omega_k') < \max\,[r(f \mid \omega_k), r(f \mid \omega_k')]. \tag{26}$$

Clearly, $f(v)$ cannot be uniformly better than $f^0(v)$ in Ω. Again, from (26),

$$r(f^0 \mid \omega_k) < \sup_{\omega \in \Omega}\,[r(f \mid \omega)], \tag{27}$$

so that, since k is arbitrary,

$$\sup_{\omega \in \Omega}[r(f^0 \mid \omega)] = \lim_{k \to \infty} r(f^0 \mid \omega_k) \leq \sup_{\omega \in \Omega}\,[r(f \mid \omega)]. \tag{28}$$

Since $f(v) \not\equiv f^0(v)$ in the preceding argument is arbitrary, we have shown that (a) no $f(v)$ can be uniformly better than $f^0(v)$ and (b) $\sup_\omega\,[r(f^0 \mid \omega)] = \inf_f \sup_\omega\,[r(f \mid \omega)]$, i.e. that $f^0(v)$ is admissible and minimax. The last part of the theorem follows upon setting $\omega_k = \omega_0$ in (27). This completes the proof of Theorem 3.

The conditions on Ω for $f^0(v)$ to be the unique minimax decision function may be written as follows:

There exist $\omega_0 = (m_1, m_2 : \sigma)$, $\omega_0' = (\mu_1, \mu_2 : \sigma)$ *in* Ω *such that*

$$\text{(i)}\ r(f^0 \mid \omega_0)(= r(f^0 \mid \omega_0')) = \sup_{\omega \in \Omega}\,[r(f^0 \mid \omega)] \qquad (\neq 0, \infty),$$

$$\text{(29)} \quad \text{(ii)}\ \mu_1 = m_2 + \left(\frac{n_1 - n_2}{n_1 + n_2}\right)(m_1 - m_2), \qquad \mu_2 = m_1 + \left(\frac{n_1 - n_2}{n_1 + n_2}\right)(m_1 - m_2),$$

$$\text{(iii)}\ h(\omega_0) = h(\omega_0').$$

For the important risk functions (1) and (6), (29)(ii) implies (29)(iii) (i.e. $h(\omega)$ depends on $|m_1 - m_2|$ alone). Moreover, when $n_1 = n_2$, (29)(ii) becomes $\mu_1 = m_2$, $\mu_2 = m_1$. Thus for (1) and (6), when $n_1 = n_2$ the conditions (29) reduce simply to the condition that *at least two points in* Ω *at which the risk for* $f^0(v)$ *is maximum be image points of one another in the plane* $\{\omega : m_1 = m_2\}$. In particular, it follows that if $n_1 = n_2$ and if the given set Ω is "symmetric" in the sense that whenever $(m_1, m_2 : \sigma)$ is in Ω then $(m_2, m_1 : \sigma)$ is also in Ω, then $f^0(v)$ is the unique minimax

decision function provided that it attains its maximum risk in Ω, the risk function in question beging (1) or (6) There are obvious modifications (involving two sequences of points in Ω) of these remarks which assert that $f^0(v)$ is at least an admissible minimax decision function in case $f^0(v)$ does not attain its maximum risk in Ω.

We shall now state the result analogous to Theorem 3 for the case when one of the means is known exactly, say $m_2 = c$. The decision function $f_c^0(v)$ is defined by (3).

THEOREM 4. *Suppose that there exist sequences* $\{\omega_k\}$, $\{\omega_k'\}$ *of points* $\omega_k = (c + a_k, c{:}\ \sigma_k)$, $\omega_k' = (c - a_k, c{:}\ \sigma_k)$ *in* Ω *such that*

(i) $\lim_{k\to\infty} r(f_c^0 \mid \omega_k) = \sup_{\omega\in\Omega} [r(f_c^0 \mid \omega)]. \qquad (\neq 0, \infty)$

(ii) $h(\omega_k) = h(\omega_k')$ *for every* $k = 1, 2, \cdots$.

Then $f_c^0(v)$ *is an admissible minimax decision function. If there exist* $\omega_0 = (c + a, c{:}\ \sigma)$, $\omega_0' = (c - a, c{:}\ \sigma)$ *in* Ω *satisfying* (i) *and* (ii), *then* $f_c^0(v)$ *is the unique minimax decision function.*

The proof (based on Theorem 2(iii)) is similar to that of Theorem 3 and will be omitted. Note that for the risk functions (1) and (6), condition (ii) is automatically satisfied.

The reader will have observed that results which may be obtained from Theorem 2(iii) in the manner of Theorems 3 and 4 will assert the optimal character of decision functions which are characteristic functions of sets of the type $\{v{:}\ a\bar{x}_1 + b\bar{x}_2 > c\}$. The following example, cited as Example (iii) of Section 1, shows that for arbitrary Ω the optimum decision function need not be of this type.

Suppose that $n_1 = n_2 = n$, that $\bar{\Omega}$ consists of the three points

$$\omega_0 = (\tfrac{1}{2}, -\tfrac{1}{2}{:}\ 1),\ \omega_1 = (\tfrac{1}{2}, \tfrac{3}{2}{:}\ 1),\ \omega_2 = (-\tfrac{3}{2}, -\tfrac{1}{2}{:}\ 1),$$

and that the risk function under consideration is given by (1) or (6). Then the unique minimax decision function is $f^{**}(v)$ given by (4), where $\lambda > 0$ is determined by

$$E[1 - f^{**} \mid \omega_0] = E[f^{**} \mid \omega_1]. \tag{30}$$

The proof follows. $f^{**}(v)$ is the characteristic function of the set $\{v{:}\ \phi(v) > c_1\phi_1(v) + c_2\phi_2(v)\}$, where ϕ, ϕ_1, ϕ_2 are the frequency functions of the probability distributions in E_{2n} corresponding to the parameter points ω_0, ω_1, ω_2 respectively, with $c_1 = c_2 = e^n/\lambda$. Since for all $\lambda > 0$,

$$E[f^{**} \mid \omega_1] = E[f^{**} \mid \omega_2],$$

and since a unique $\lambda > 0$ satisfying (30) certainly exists, it follows (cf. (19) and (C)) that

$$r(f^{**} \mid \omega_0) = r(f^{**} \mid \omega_1) = r(f^{**} \mid \omega_2) = B,$$

say. Let $f(v)$ be any decision function $\not\equiv f^{**}(v)$. We shall show that

$$B < \max\,[r(f \mid \omega_0), r(f \mid \omega_1), r(f \mid \omega_2)]. \tag{31}$$

Suppose not. Then

$$r(f \mid \omega_1) = E[f \mid \omega_1] \leq E[f^{**} \mid \omega_1] = r(f^{**} \mid \omega_1),$$
$$r(f \mid \omega_2) = E[f \mid \omega_2] \leq E[f^{**} \mid \omega_2] = r(f^{**} \mid \omega_2).$$

Then, by Theorem 1, Note 2, we must have $E[f \mid \omega_0] < E[f^{**} \mid \omega_0]$, so that

$$r(f \mid \omega_0) = 1 - E[f \mid \omega_0] > 1 - E[f^{**} \mid \omega_0] = r(f^{**} \mid \omega_0) = B,$$

contrary to hypothesis. Hence (31) holds, and since $f(v) \not\equiv f^{**}(v)$ is arbitrary our assertion is proved. (Note that

$$r(f^0 \mid \omega_0) = r(f^0 \mid \omega_1) = r(f^0 \mid \omega_2)$$

also, so that $f^{**}(v)$ is uniformly better than $f^0(v)$ in $\bar{\Omega}$.) We remind the reader that $f^{**}(v)$ remains the unique minimax decision function with respect to (1) or (6) and any Ω which contains ω_0, ω_1, ω_2, and is such that $\sup_{\omega \in \Omega} [r(f^{**} \mid \omega)] = B$. Whether a set Ω satisfies the last condition will in general depend on whether the risk function in question is (1) or (6).

3. Examples and discussion. In this section we shall discuss the relevance of Theorems 3 and 4 to two specific problems of the greater mean. The examples given are purely illustrative and the reader will readily construct others in which the statistician is faced with similar problems of decision.

Example 1. A farmer F has tested two varieties π_1, π_2 of grain in a field experiment in which n_i plots were assigned to π_i, $i = 1, 2$, all plots being of equal area. The plot yields obtained were $y_{11}, y_{12}, \cdots, y_{1n_1}$ and $y_{21}, y_{22}, \cdots, y_{2n_2}$ bushels respectively F gives this data to a statistician S for analysis. F is willing to assume that the yields per plot for each of the two varieties are normally distributed with unknown means μ_1, μ_2 and a common variance, also unknown. F says he is particularly interested in whether the two varieties are "significantly different."

S is well aware that F's interest in the varieties is not purely scientific—that is to say, F did not perform the field experiment for the sole purpose of estimating the unknown parameters or testing hypotheses concerning them. S also knows that it is very unlikely that μ_1 *is* equal to μ_2.

Suppose that in fact F wishes to decide which variety he should use next year on his land in order to make the maximum possible profit, and is afraid that if he were to act as if the observed mean yields $\bar{y}_1$, $\bar{y}_2$ were the true population mean yields, he might make a gross error. So F is willing to compromise between the two varieties (that is, he will assign some fraction f of his land to π_1 and the rest to π_2) in case S declares that there is no evidence of the two varieties being different.

If this is the case, S should ask F how much it costs him to use π_i and the price at which he expects to sell his grain. Supposing that these quantities are a_i dollars per acre and b dollars per bushel respectively, and that the area of each plot in the field experiment was c acres, S will set

$$m_i = \text{expected profit per acre in using variety } \pi_i$$
$$= (b/c)\mu_i - a_i \quad \text{dollars} \qquad (i = 1, 2),$$
$$\omega = (m_1, m_2 : \sigma),\ \sigma^2 \text{ being the variance of the profit per acre in using } \pi_i \qquad (i = 1, 2),$$
$$\gamma(\omega) = (m_1, m_2) \qquad \text{(see Section 2, (C))},$$
$$x_{ij} = (b/c)y_{ij} - a_i,\ \bar{x}_i = n_i^{-1} \cdot \sum_{j=1}^{n_i} x_{ij}, \qquad v = (x_{11}, \cdots, x_{1n_1}; x_{21}, \cdots, x_{2n_2}),$$

so that $r(f \mid \omega)$ is given by (1) and is equal to the expected loss (in terms of profit per acre) incurred by using the proportions $f(v)$, $1 - f(v)$ of the varieties π_1, π_2 as compared with using the variety with the greater mean for the whole of the land Then if S is satisfied that the set Ω of possible points ω satisfies the conditions of Theorem 3 he should recommend that F use π_1 alone if $\bar{x}_1 > \bar{x}_2$, and π_2 alone if $\bar{x}_2 > \bar{x}_1$, this being the safest procedure in the sense that it is the minimax strategy (cf. Example 1 in [3]).

We shall illustrate by a simple example the obvious method of verifying whether $f^0(v)$ is the minimax decision function for a given Ω. We have by (22), using the risk function (1) obtained by setting $\gamma(\omega) = (m_1, m_2)$,

$$r(f^0 \mid \omega) = h(\omega)G(-\mid d(\omega) \mid)$$
$$= \mid m_1 - m_2 \mid G\left(-\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} \mid m_1 - m_2 \mid /\sigma\right). \tag{32}$$

Now suppose that

$$\Omega = \left\{\omega : a - \frac{l}{2} \le m_1 \le a + \frac{l}{2},\ b - \frac{l}{2} \le m_2 \le b + \frac{l}{2} : \sigma_0 - \rho \le \sigma \le \sigma_0\right\}, \qquad l > \mid a - b \mid, \tag{33}$$

where $a, b, l, \sigma_0, \rho(\ge 0)$ are certain constants. By (32), the maximum risk occurs at some points in Ω for which $\sigma = \sigma_0$. We have

$$r(f^0 \mid \sigma = \sigma_0) = \sigma_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}} \cdot [xG(-x)], \tag{34}$$

where

$$x = x(\omega) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} \mid m_1 - m_2 \mid /\sigma_0 .$$

If $a = b$ and $n_1 = n_2$ we see from the remark following (29) that $f^0(v)$ is the unique minimax decision function. Suppose therefore that $a \neq b$ or $n_1 \neq n_2$ or both. Now

$$\sup_x [xG(-x)] = x_0 G(-x_0) = .1700 \text{ (approx.)}, \tag{35}$$

where $x_0 = 7518$ (approx.). If m_1, m_2 were unrestricted, $r(f^0 \mid \sigma = \sigma_0)$ would be a maximum when $|m_1 - m_2| = \sigma_0 x_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}$, by (34) and (35). Hence $f^0(v)$ will be the unique minimax decision function if these two lines intersect the square $\left\{a - \frac{l}{2} \leq m_1 \leq a + \frac{l}{2}, b - \frac{l}{2} \leq m_2 \leq b + \frac{l}{2}\right\}$ in such a way that at least two points lying on these lines and in the square satisfy (29)(ii). This will be the case if

$$l > \max\left[\,|a - b| + y_0\,,\ \max(\,|a - b|\,,\, y_0) + \left|\frac{n_1 - n_2}{n_1 + n_2}\right| y_0\right], \tag{36}$$

where

$$y_0 = x_0 \sigma_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}.$$

We have assumed that $l > |a - b|$, for otherwise either $m_1 \leq m_2$ or $m_1 \geq m_2$ for all ω in Ω, and there is no problem. It is therefore clear that for n_1 and n_2 sufficiently large, $f^0(v)$ will be the unique minimax decision function. That (36) is not a very strong requirement may be seen by setting $a = b$, $n_1 = 2n_2$, in which case (36) reduces to

$$l > \sigma_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2} \qquad \text{(approx.)}.$$

We remark that $f^0(v)$ remains the unique minimax decision function for any n_1, n_2 "when $l = \infty$" so that Ω is given by

$$\Omega = \{\omega: -\infty < m_1 < \infty, -\infty < m_2 < \infty : \sigma_0 - \rho \leq \sigma \leq \sigma_0\}. \tag{33'}$$

It is of interest to consider the "one sample" case when one of the means is known, say $m_2 = c$. This will be the case (approximately) if π_2 is a standard variety which has been in use for some time and π_1 is a new variety. The analogue of the parameter space discussed above is then

$$\Omega = \left\{\omega: m_2 = c, a - \frac{l}{2} \leq m_1 \leq a + \frac{l}{2} : \sigma_0 - \rho \leq \sigma \leq \sigma_0\right\}, \quad \frac{l}{2} > |a - c|. \tag{37}$$

By using Theorem 4 it can be seen that $f_c^0(v)$ as defined by (3) is the unique minimax decision function if $c = a$ or if c is not necessarily equal to a, but

$$\frac{l}{2} - |a - c| > \sigma_0 x_0 \left(\frac{1}{n_1}\right)^{1/2}, \tag{38}$$

where x_0 is given by (35). Since the left-hand side of (38) is positive, it is clear that $f_c^0(v)$ will be the unique minimax decision function with respect to (37) if

n_1 is sufficiently large. Note that $f_c^0(v)$ is the unique minimax decision function for any n_1 when $l = \infty$ and Ω is given by

$$\Omega = \{\omega: m_2 = c, -\infty < m_1 < \infty : \sigma_0 - \rho \leq \sigma \leq \sigma_0\}. \tag{37'}$$

The reader may find it instructive to consider other plausible sets Ω which satisfy the conditions of Theorems 3 and 4 and also some which do not, assuming $\sigma = 1$ for simplicity. It should be observed that no matter what Ω may be, provided only that $\sigma \leq \sigma_0$ for all ω in Ω, we shall have by (32) and (35)

$$\sup_{\omega \in \Omega} [r(f^0 \mid \omega)] \leq .1700 \cdot \sigma_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}} \qquad \text{(approx.)}.$$

In a similar way it can be seen that for any Ω in which m_2 equals c and $\sigma \leq \sigma_0$,

$$\sup_{\omega \in \Omega} [r(f_c^0 \mid \omega)] \leq .1700 \cdot \sigma_0 \cdot \left(\frac{1}{n_1}\right)^{\frac{1}{2}} \qquad \text{(approx)}$$

EXAMPLE 2. π_1 and π_2 are two soporific drugs, the random variables generated by them being the duration of sleep induced by a standard dose in an individual chosen at random. It is assumed that these two populations are normal with unknown means m_1, m_2 and a common variance σ^2, also unknown. In a series of independent trials in which n_1 individuals received the first drug and n_2 the second, the outcome was $v = (x_{11}, x_{12}, \cdots, x_{1n_1}, x_{21}, x_{22}, \cdots, x_{2n_2})$. The statistician S is required to say which is the more effective drug

Here a reasonable risk function is (6), where $f(v)$ takes on only the values 0, 1, corresponding to the decisions "$m_1 \leq m_2$" and "$m_1 \geq m_2$" respectively.[4] The problem of choosing $f(v)$ so as to minimize this risk was considered by Simon [4]. He showed that in case $n_1 = n_2$, $f^0(v)$ is the *uniformly best* decision function in the class of symmetric decision functions. (Given $n_1 = n_2 = n$, a decision function $f(v)$ is said to be symmetric if $f(x_{11}, x_{12}, \cdots, x_{1n}; x_{21}, x_{22}, \cdots, x_{2n}) \equiv 1 - f(x_{21}, x_{22}, \cdots, x_{2n}; x_{11}, x_{12}, \cdots, x_{1n})$. See also [3].) It is natural to confine oneself to the class of symmetric decision functions when the sample sizes are equal, but under the implicit assumption that if $\omega = (a, b: \sigma)$ is a possible parameter point, then $\omega' = (b, a: \sigma)$ is also (cf. the remarks following (29)). The illustrations in Section 1 show that if the sample sizes are unequal or if Ω is not symmetric in the sense just described, there may exist decision functions which are *uniformly better* than $f^0(v)$: in (i) we have a "symmetric" Ω but $n_1 \neq n_2$, in (iii), $n_1 = n_2$ but Ω is not "symmetric."

However, $f^0(v)$ is an admissible minimax decision function no matter what the sample sizes, provided only that Ω satisfies a certain not too restrictive condition. We have

$$\bar{r}(f^0 \mid \omega) = \begin{cases} G(-\mid d(\omega) \mid) & \text{for } m_1 \neq m_2, \\ 0 & \text{for } m_1 = m_2. \end{cases} \tag{39}$$

[4] For some purposes it would be more appropriate to take (1) as the risk function for this problem, letting the decision functions $f(v)$ take on only the values 0 and 1 We have (essentially) discussed this case in the previous example.

It is clear that if $\{\omega_k\}$ is a sequence of points in Ω such that

$$\lim_{k\to\infty} d(\omega_k) = 0, \quad \text{then} \quad \lim_{k\to\infty} \bar{r}(f^0 \mid \omega_k) = \frac{1}{2} = \sup_{\omega\epsilon\Omega} [\bar{r}(f^0 \mid \omega)].$$

Therefore, by Theorem 3, $f^0(v)$ is admissible and minimax if some point in the plane $\{\omega: m_1 = m_2\}$ is an interior point of the set Ω of possible parameter points (in fact it is sufficient if some plane $\sigma = \sigma_0(>0)$ intersects Ω in a set which has an interior point on the line $m_1 = m_2$). Hence if nothing much is known about the two drugs, S could regard the foregoing as a justification for asserting "$m_1 \geq m_2$" if $\bar{x}_1 > \bar{x}_2$ and "$m_1 \leq m_2$" otherwise.

We have given no criterion for the choice of a suitable decision function when two or more admissible minimax decision functions exist, and our diffidence in recommending the use of $f^0(v)$ in the present case is due to the fact that under the condition stated above there will exist decision functions other than $f^0(v)$ which are also admissible and minimax with respect to (6). Let us suppose that Ω is given by (33). Then $f^0(v)$ is admissible and minimax, by the preceding paragraph. However, it follows from Theorem 4 that each of

$$f^0_{c_1}(v) = \begin{cases} 1 \text{ if } \bar{x}_1 > c_1, \\ 0 \text{ otherwise,} \end{cases} \quad \text{and} \quad f^{c_2}_0(v) = \begin{cases} 0 \text{ if } \bar{x}_2 > c_2, \\ 1 \text{ otherwise,} \end{cases}$$

is also admissible and minimax, where c_1 and c_2 are arbitrary constants with $\max [a, b] - \frac{l}{2} \leq c_1, c_2 \leq \min [a, b] + \frac{l}{2}$.

There is, however, some reason for preferring $f^0(v)$ to other decision functions in the present case. S has been asked to give his opinion as to which is the better drug, and presumably no immediate consequences follow from the opinion which he might express. (This would not be the case if there were a sleepless individual on hand who had to be given a dose of one of the two drugs Cf. footnote 4.) Although the problem *is* of a scientific nature, insistence upon literal exactitude in the interpretation of "incorrect decision" is meaningful only insofar as it is compatible with the physical situation. In view of the limited determinacy of unknown parameters in general, and of the limitations of experiments on soporific drugs in particular, it may be possible and even desirable to modify (6) in such a way that for any fixed σ the risk tends to zero with $| m_1 - m_2 |$. Thus modified, the risk function would be essentially similar to (1). A rather drastic way of introducing this modification would be to agree that the assertion of equality of the two means does not constitute an error in case $| m_1 - m_2 | < \epsilon$, where ϵ is some positive constant. S will then take

$$(40) \qquad \bar{r}_\epsilon(f \mid \omega) = \begin{cases} \bar{r}(f \mid \omega) \text{ if } | m_1 - m_2 | \geq \epsilon, \\ 0 \text{ otherwise,} \end{cases}$$

as the risk function. (Note that in using $\bar{r}_\epsilon(f \mid \omega)$ rather than $\bar{r}(f \mid \omega)$, S has in effect deleted the set $\{\omega: | m_1 - m_2 | < \epsilon\}$ from the given set Ω by defining $\gamma(\omega) =$

(0, 0) there, instead of only when $m_1 = m_2$ as in the case of $\bar{r}(f \mid \omega)$. Cf "zones of indifference," [5, pp 27–30]). It follows from Theorem 3 that $f^0(v)$ is the unique minimax decision function with respect to (40) and (33) if $a = b$ and $n_1 = n_2$ and also if at least one of these conditions does not hold but

$$l > \max \left[\mid a - b \mid + \epsilon, \max \left(\mid a - b \mid, \epsilon \right) + \left| \frac{n_1 - n_2}{n_1 + n_2} \right| \epsilon \right].$$

Thus $f^0(v)$ will be the unique minimax decision function no matter what n_1, n_2, a, b or l may be, provided only that ϵ is sufficiently small. We shall leave other modifications of $\bar{r}(f \mid \omega)$ and discussion of $\bar{r}(f \mid \omega)$ with respect to other types of parameter spaces (e.g. (37)) to the reader.

We conclude this discussion with a remark on the proper choice of n_1 and n_2 in using $f^0(v)$ when the risk function belongs to the class R defined in Section 2, (C). (The risk functions (1) and (6) belong to R) Suppose that before experimentation starts, it is agreed that one must have $n_1 + n_2 = 2k$, where k is a fixed integer. In that case, choosing $n_1 = n_2 = k$ will be the best choice of n_1, n_2 in the following sense. (a) For any fixed ω, $r(f^0 \mid \omega)$, which is the expected loss, then becomes a minimum. This follows immediately from (22), since

$$r(f^0 \mid \omega) = h(\omega)G(- \mid d(\omega) \mid), \qquad \mid d(\omega) \mid = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} \mid m_1 - m_2 \mid / \sigma,$$

and $\mid d(\omega) \mid$ has its maximum when $n_1 = n_2 = k$. (b) For any fixed ω, the *variance* of the loss also becomes a minimum. In using $f^0(v)$, the loss takes the values 0 and $h(\omega)$ only, with $P(\text{loss} = h(\omega) \mid \omega) = G(- \mid d(\omega) \mid) = \alpha$ say. Therefore, the variance of the loss is $h^2\alpha(1 - \alpha)$. Since $\alpha \leq \frac{1}{2}$, this expression increases with increasing α, and so has its minimum when $n_1 = n_2 = k$. This remark is, of course, without prejudice to the question of whether $f^0(v)$ is admissible and minimax with respect to a given Ω for every n_1 and n_2 with $n_1 + n_2 = 2k$

4. A remark on randomized decision functions. In the foregoing discussion we have confined attention to the class of non-randomized decision functions: the space of possible decisions being some subset of $0 \leq f \leq 1$, the statistician constructs (in advance) a suitable decision function $f(v)$, obtains a particular sample point v by sampling the two populations, and takes $f(v)$ as his decision. It is, however, of some theoretical interest to consider more general formulations in which the decision arrived at by the statistician may be a random function of the sample point v.

A randomized decision function can be defined in several ways One definition is as follows. Let $\phi(z \mid v)$ be a function defined for all v in E_N and all real z such that for any fixed z it is a measurable function of v, and such that for any fixed v it is the distribution function of a random variable with values in $0 \leq z \leq 1$. We shall denote this random variable by $Z_\phi(v)$ and call it a (randomized) decision function. In using it, the statistician first obtains a particular point v by sampling the two populations, then performs a random experiment whose outcome Z

has the known distribution function $P(Z \leq z) = \phi(z \mid v)$, and takes Z as his decision. The class of all decision functions corresponding to all functions $\phi(z \mid v)$ will be denoted by $\{Z_\phi(v)\}$. It is clear that this class includes the class of non-randomized decision functions.

This definition of the structure of randomized decision functions follows the method described by Halmos and Savage in their interesting remarks ([6], pp. 239–241) on the value of sufficient statistics in statistical methodology. For any $Z_\phi(v)$, we have

$$
\begin{aligned}
P(Z_\phi(v) \leq z \mid \omega) &= \int_{E_N} P(Z_\phi(v) \leq z \mid \omega, v)\, dK(v \mid \omega) \\
&= \int_{E_N} \phi(z \mid v)\, dK(v \mid \omega).
\end{aligned} \tag{41}
$$

We shall now show that in all problems of the greater mean in which the methods of Section 2 can be applied to non-randomized decision functions, randomization cannot be recommended. More precisely, the following holds.

THEOREM. *Let $\bar{f}(v)$ be a non-randomized decision function which takes on only the values 0 and 1 and which is the unique non-randomized decision function whose expected value $E[\bar{f} \mid \omega]$ satisfies a certain condition Q as a function of ω. Then $\bar{f}(v)$ is the unique decision function whose expected value satisfies the condition Q;* i.e. if *$Z_\phi(v)$ is a decision function such that $E[Z_\phi \mid \omega]$ satisfies Q, then*

$$
P(\bar{f}(v) = Z_\phi(v) \mid \omega) = 1 \quad \textit{for all } \omega. \tag{42}
$$

It follows in particular that Theorem 2 remains valid with the arbitrary non-randomized $f(v)$ replaced by an arbitrary $Z_\phi(v)$, and in consequence, Theorems 3 and 4 remain valid when the class of decision functions in question is $\{Z_\phi(v)\}$.

PROOF. Let $Z_\phi(v)$ be a decision function whose expected value satisfies the condition Q. Now, by (41) and Theorem 5 of [7] we have

$$
E[Z_\phi \mid \omega] = \int_{E_N} f^\phi(v)\, dK(v \mid \omega) = E[f^\phi \mid \omega], \tag{43}
$$

where

$$
f^\phi(v) = \int_0^1 z d_z\phi(z \mid v), \qquad 0 \leq f^\phi(v) \leq 1. \tag{44}
$$

It is clear from (43) that $E[f^\phi \mid \omega]$ satisfies Q and so we must have

$$
f^\phi(v) = \bar{f}(v) \text{ a.e.} \tag{45}
$$

by hypothesis. Since $\bar{f}(v)$ takes on only the values 0 and 1, it follows from (44) and (45) that

$$
\int_{\{z=\bar{f}(v)\}} d_z\phi(z \mid v) = 1 \text{ a.e.},
$$

which implies (42). In order to verify the last part of the remark, consider any particular problem of the greater mean The risk function of any decision function $Z_\phi(v)$ is, by (15),

$$r(Z_\phi \mid \omega) = W(\omega, E[Z_\phi \mid \omega]).$$

Hence a condition on the risk function of Z_ϕ is equivalent to a condition on $E[Z_\phi \mid \omega]$ as a function of ω, and the truth of the remark follows by appropriate definition of the condition Q in terms of the risk function.

REFERENCES

[1] A. WALD, "Statistical decision functions," *Annals of Math Stat*, Vol 20 (1949), pp 165–205.

[2] J NEYMAN AND E S PEARSON, "Contributions to the theory of testing statistical hypotheses," *Stat Res Memoirs*, Vol. I (1936), pp. 1–37.

[3] R. R. BAHADUR, "On a problem in the theory of k populations," *Annals of Math. Stat*, Vol. 21 (1950), pp. 362–375

[4] H A SIMON, "Symmetric tests of the hypothesis that the mean of one normal population exceeds that of another," *Annals of Math Stat*, Vol 14 (1943), pp. 149–154.

[5] A. WALD, *Sequential analysis*, John Wiley and Co., 1947

[6] P. R. HALMOS AND L. J SAVAGE, "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Annals of Math Stat*, Vol 20 (1949), pp. 225–241

[7] H ROBBINS, "Mixture of distributions," *Annals of Math Stat.*, Vol 19 (1948), pp 360–369.

ANALYSIS OF EXTREME VALUES

By W. J. Dixon[1]

University of Oregon

1. Introduction. It is well recognized by those who collect or analyze data that values occur in a sample of n observations which are so far removed from the remaining values that the analyst is not willing to believe that these values have come from the same population. Many times values occur which are "dubious" in the eyes of the analyst and he feels that he should make a decision as to whether to accept or reject these values as part of his sample. On the other hand he may not be looking for an error, but may wish to recognize a situation when an occasional observation occurs which is from a different population. He may wish to discover whether a significant analysis of variance indicates an extreme value significantly different from the remainder. Also, of course, the extreme value may differ significantly without causing a significant analysis of variance and he may wish to discover this. It is reasonable to suppose that a criterion for rejecting observations would be useful here also. The choice of a suitable criterion for rejecting observations introduces a number of questions.

1. Should any observations be removed if we wish a representative sample including whatever contamination arises naturally? In other words, it may be desirable to describe the population including *all* observations, for only in that way do we describe what is actually happening.

2. If the analyst wishes to sample the population unaffected by contamination he must either remove the contaminating items or employ statistical procedures which reduce to a minimum the effect of the contamination on the estimates of the population. That is, he may wish to describe only 95% of his population if the description is altered radically by the remaining 5% of the observations. He may have external reasons which are good and sufficient for wishing to describe only 95% of his observations. Suppose he wishes to use the sample for a statistical inference; the inclusion of all the data may sufficiently violate the assumptions underlying the inference to exclude the possibility of making a valid inference

This paper will concern itself only with those problems which arise from Question 2.

If we wish to follow some procedure which attempts to remove contamination we must consider the performance of any proposed criterion with respect to the proportion of contamination the criterion will discover and, of course, the proportion of the "good" observations which are removed by the use of the criterion. But, perhaps more important, we must consider what sort of bias will result when the standard statistical procedures are applied to samples of observations which have been processed in this manner.

[1] This paper was prepared under a contract with the Office of Naval Research

If we wish to follow a procedure which will not search for particular values to be excluded but will minimize their effect if present, we must investigate the sampling distributions of these modified statistics and estimate the loss in information resulting from their use when all observations are "good." We must also investigate the expected bias which will result when "bad" items are present even though essentially excluded. Perhaps most disturbing about the avoidance of "bad" items is the fact that a decision must still be made as to whether a "bad" item was present or not in order to know in which way our estimates may be biased For example, a sample mean computed by avoiding the two end observations will not be a biased estimate of the mean of a symmetric population if both end items should actually be included or if both end items should not be included. However, if only one of the two should not be included this estimate of the mean will be biased

2. Models of contamination. The performance of the various criteria for discovery of one or more contaminators will be measured with reference to contaminations of the following two types entering into samples of observations from a normal population with mean μ and variance σ^2, $N(\mu, \sigma^2)$

A. One or more observations from $N(\mu + \lambda\sigma, \sigma^2)$,

B One or more observations from $N(\mu, \lambda^2\sigma^2)$.

A represents the occurrence of an "error" in mean value such as will occur in dial readings when errors are made in reading incorrectly digits other than the last one or two digits. Errors of this sort may result from momentary shifts in line voltage or from the inclusion among a group of objects of one or two items of completely different origin. This type of contamination will be referred to as "location error." B represents the occurrence of an "error" from a population with the same mean but with a greater variance than the remainder of the sample. This type of error will be referred to as a "scalar error." It is likely that many errors could be better described as a combination of A and B, but a study of these two errors separately should throw considerable light on the question of "gross errors" or "blunders."

Many authors have written on the subject of the rejection of outlying observations. Apparently none have been successful in obtaining a general solution to the problem Nor has there been success in the development of a criterion for discovery of outliers by means of a general statistical theory; e g., maximum likelihood. A large number of criteria have been advanced on more or less intuitive grounds as appropriate criteria for this purpose. In no case was investigation made of the performance of these criteria except for a few illustrative examples.

References for the criteria discussed in the next section are given at the end of this paper. Indications are given as to the significance values available in those papers

3. Criteria to be considered. The performance of two types of criteria has been investigated for samples contaminated with location or scalar errors.

a) σ known or estimated independently,

b) σ unknown.

The n observations are ordered $x_1 < x_2 < \cdots < x_n$. The criteria involving external knowledge of σ are:

A. χ^2 test,

$$\chi^2 = \frac{\Sigma(x - \bar{x})^2}{\sigma^2}.$$

B. Extreme deviation,

$$B_1 = \frac{x_n - \bar{x}}{\sigma} \left(\text{or } \frac{\bar{x} - x_1}{\sigma}\right),$$

$$B_2 = \frac{x_n - x_{n-1}}{\sigma} \left(\text{or } \frac{x_2 - x_1}{\sigma}\right).$$

C. Range,

$$C_1 = \frac{w}{\sigma}, \qquad w = x_n - x_1,$$

$$C_2 = \frac{w}{s}, \quad s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} \qquad (s \text{ independently estimated}).$$

The criteria involving only the information of a single sample of n observations are:

D. Modified F test.

1. For single outlier x_1,

$$D_1 = \frac{S_1^2}{S^2}, \quad \text{where} \quad S_1^2 = \sum_2^n (x - \bar{x}_1)^2, \quad \bar{x}_1 = \sum_2^n x/(n - 1),$$

$$S^2 = \sum_1^n (x - \bar{x})^2, \quad \bar{x} = \sum_1^n x/n$$

$$\left(\text{or for } x_n,\ D_1 = \frac{S_n^2}{S^2}\right).$$

2. For double outliers x_1, x_2,

$$D_2 = \frac{S_{1,2}^2}{S^2}, \quad \text{where} \quad S_{1,2}^2 = \sum_3^n (x - \bar{x}_{1,2})^2, \quad \bar{x}_{1,2} = \sum_3^n x/(n-2)$$

$$\left(\text{or for } x_n, x_{n-1},\ D_2 = \frac{S_{n,n-1}^2}{S^2}\right).$$

E. Ratios of ranges and subranges.

1. For single outlier x_1,

$$r_{10} = \frac{x_2 - x_1}{x_n - x_1}$$

$$\left(\text{or for } x_n,\ r_{10} = \frac{x_n - x_{n-1}}{x_n - x_1}\right).$$

2. For single outlier x_1 avoiding x_n,

$$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$$

$$\left(\text{or for } x_n \text{ avoiding } x_1,\ r_{11} = \frac{x_n - x_{n-1}}{x_n - x_2}\right).$$

3. For single outlier x_1, avoiding x_n, x_{n-1},

$$r_{12} = \frac{x_2 - x_1}{x_{n-2} - x_1}$$

$$\left(\text{or for } x_n \text{ avoiding } x_1,\ x_2,\ r_{12} = \frac{x_n - x_{n-1}}{x_n - x_3}\right).$$

4. For outlier x_1 avoiding x_2,

$$r_{20} = \frac{x_3 - x_1}{x_n - x_1}$$

$$\left(\text{or for } x_n \text{ avoiding } x_{n-1},\ r_{20} = \frac{x_n - x_{n-2}}{x_n - x_1}\right).$$

5. For outlier x_1 avoiding x_2 and x_n,

$$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1}$$

$$\left(\text{or for } x_n \text{ avoiding } x_{n-1},\ x_1,\ r_{21} = \frac{x_n - x_{n-2}}{x_n - x_2}\right).$$

6. For outlier x_1 avoiding x_2 and x_n, x_{n-1},

$$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$$

$$\left(\text{or for } x_n \text{ avoiding } x_{n-1},\ x_1,\ x_2,\ r_{22} = \frac{x_n - x_{n-2}}{x_n - x_3}\right).$$

F. Extreme deviation and standard deviation.
For single outlier x_n,

$$F = \frac{x_n - \bar{x}}{s} \quad \left(\text{or for } x_1,\ F = \frac{\bar{x} - x_1}{s}\right).$$

The performance of the large number of criteria listed here will be assessed with respect to discovery of contamination of the type given in Section 2.

4. Performance of criteria (estimate of σ available). The χ^2 test will of course give an indication of a large dispersion and since the extreme values are chief contributors to the sum of squares, it is possible to use this test as a criterion for rejecting a value or values which are at the greatest distance from the mean It might be supposed the B_1 and B_2 would give better results since particular attention is paid to the end item. The same argument would influence one in favor of C_1 or C_2. The performance of C_2 can, of course, be expected to vary with the degrees of freedom in the independent estimate of σ. For this study the degrees of freedom for this estimate were held to the single value 9 d.f.

χ^2 may be used since if the value of χ^2 is too large (greater than some upper percentage point for χ^2) we might reject the value most distant from the mean. χ^2 tables may be used for percentage points. Percentage points for the other statistics considered here are given in the references at the end of this paper.

The criteria A, B_1, B_2, C_1, C_2 were investigated for $\alpha = 1\%$, 5% and 10% for $\lambda = 2, 3, 5, 7$, where one or more items are selected from a population $N(\mu + \lambda\sigma, \sigma^2)$ and the remainder from $N(\mu, \sigma^2)$. Investigations were also made for one item from $N(\mu, \lambda^2\sigma^2)$ for $\lambda = 2, 4, 8, 12$. The investigation was carried out by sampling methods The performances of different criteria were assessed for the same group of samples in order to obtain more precision in the comparison of the different tests. All of the points appearing on the graphs in the subsequent sections of this paper were based on from 66 to 200 determinations.

The performance of the above criteria is measured by computing the proportion of the time the contaminating distribution provides an extreme value and the test discovers this value. Of course, performance could be measured by the proportion of the time the test gives a significant value when a member of the contaminating population is present in the sample, even though not at an extreme. However, since it is assumed that discovery of an outlier will frequently be followed by the rejection of an extreme we shall consider discovery a success only when the extreme value is from the contaminating distribution.

The performance was judged by applying the criteria to each sample, always suspecting an outlier in the direction of the shifted mean for location error. Since the location errors were inserted by adding a fixed value to one or more of the observations, the largest value was tested as an outlier. The measure of performance was the percentage of location errors identified. When the location error was not an outlier, no test was performed and a failure for the test recorded.

In the case of the model of contamination involving the scalar error, the value was suspected which was farthest from the mean. This of course, alters somewhat the level of significance, but this procedure was followed alike for all criteria investigated. The performance was measured in the same fashion as for location errors.

Considering first, location errors, a study of the performance curves showing the per cent discovery of contaminators plotted against λ (the number of standard deviation units the population of contaminators is removed from the remainder), shows that the level of performance for σ known is considerably above the level

of performance when σ is not known The difference is greater for $n = 5$ than for $n = 15$ and, of course, the difference will diminish as the sample size increases. Figure 1 shows the performance curves for $\alpha = 5\%$ (5% significance level for the test for an outlier) of $B_1 = (x_n - \bar{x})/\sigma$ for $n = 5$ and $n = 15$ and of $r_{10} = \frac{x_n - x_{n-1}}{x_n - x_1}$ for $n = 5$ and $n = 15$

The graphs for $\alpha = 1\%$ and 10% would be similar in appearance. Figure 2 indicates the change in performance for $\alpha = 1\%$, 5%, and 10% The curves plotted are for $B_1 = (x_n - \bar{x})/\sigma$ The curves for A, B_2, C_1, C_2 show very similar results

The curve for test B_1 was used in Figures 1 and 2 since it gives the best performance of all criteria which are considered here if a single location error is present The curves showing the comparative performance of these criteria as

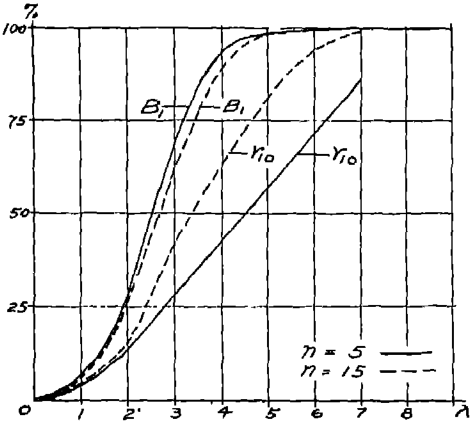


FIG. 1. Improvement in performance obtained with knowledge of σ, $\alpha = 5\%$, $n = 5$, 15

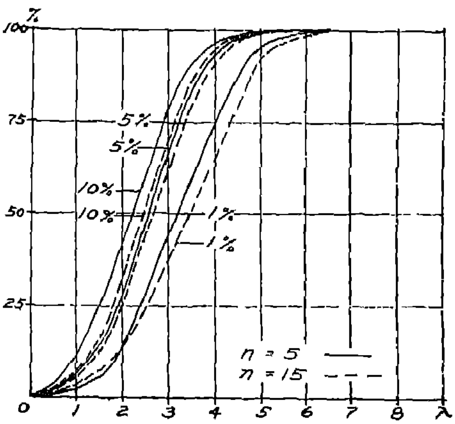


FIG. 2 The effect of the level of significance on the performance of B_1; $\alpha = 1\%$, 5%, 10%; $n = 5$, 15

well as one to be considered later (r_{10}) are given in Figure 3 for $\alpha = 5\%$ and for $n = 5$ and $n = 15$.

The following statements can be made from inspection of Figure 3:

a) The differences among A, B_1, B_2, and C_1 are not great.

b) The knowledge of σ is less important in larger samples.

c) The curve for C_2 lies above that of r_{10} for $n = 5$ and below that of r_{10} for $n = 15$. This is consistent with the use of 9 d f in the independent estimate of σ

If the question of ease in computation or application is important, it may be desirable to use B_2 or C_1 in place of B_1 for they are slightly easier to compute and it is not necessary to measure all observations to obtain the value of these statistics From Figure 3 it will be noted that the performances of these criteria are nearly as good as for B_1. If two outliers may be expected in a single sample,

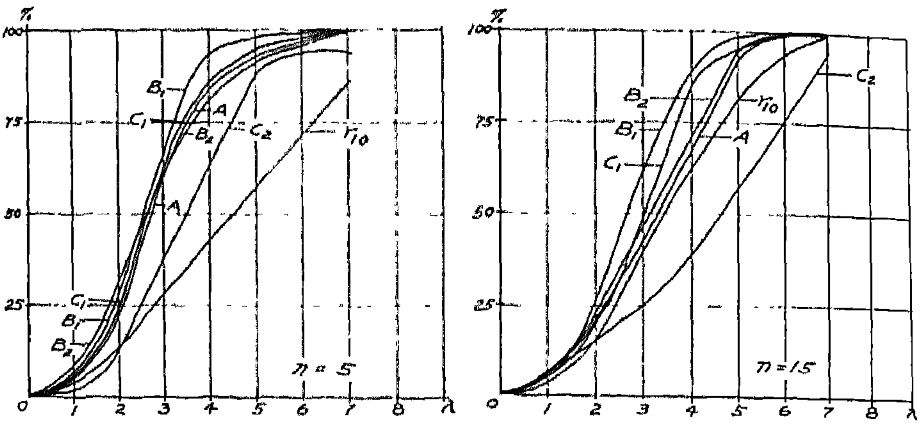


FIG 3. Comparison of the performance of criteria using σ known (or using external estimates of σ) and r_{10} for samples of size 5 and 15, $\alpha = 5\%$.

the performance of B_2 will be lowered and the performance of B_1 and C_1 will be improved. Any differences between the performance of B_1 and the performance of C_1 when two outliers are present was not discernable for $n = 5$ or 15. Figure 4 illustrates the improvement in performance for B_1 for $\alpha = 5\%$ and $n = 15$.

The performance curves of these criteria if a scalar error is present are very similar to those above except that:

1. A high level of performance is approached very slowly. For example, see Figure 5 showing the performance of B_1 and r_{10} for $n = 5$ and $n = 15$ and $\alpha = 5\%$.

2. There is a smaller difference in the performance between the criteria with σ known and σ unknown (see Figure 5).

The performance of B_1 and C_1 are noticeably increased by the introduction of more contaminators while that of B_2 decreases. No difference in the perform-

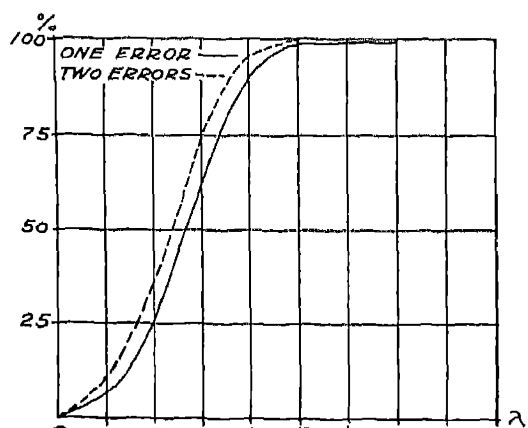


FIG. 4 Comparison of the performance of B_1 for one and two location errors in samples of size 15, $\alpha = 5\%$

ance of B_1 and C_1 were noted for either $n = 5$ or $n = 15$. Figure 6 shows the increase in performance of two contaminators for B_1 for $n = 15$, $\alpha = 5\%$.

The general recommendations for possibilities of either type of contamination, location or scalar errors, would lead one to the use of B_1 or C_1 if σ is known.

Criterion C_1 is recommended since:

1 Its performance is almost as good as the performance of B_1 for a single outlier. Their performances are about equal for two outliers and C_1 affords protection for outliers either above or below the mean.

2. It is simple to compute.

If ease of computation is not essential and maximum performance is desired, the criterion B_1 should be used. The performance of C_2 will approach that of B_1 as the number of degrees of freedom in the denominator increases.

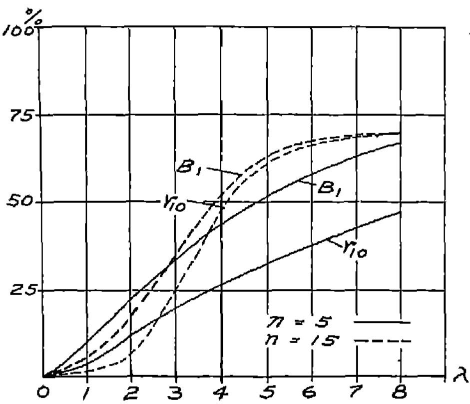


FIG. 5 Comparison of the performance of B_1 and r_{10} for one scalar error for samples of size 5 and 15, $\alpha = 5\%$

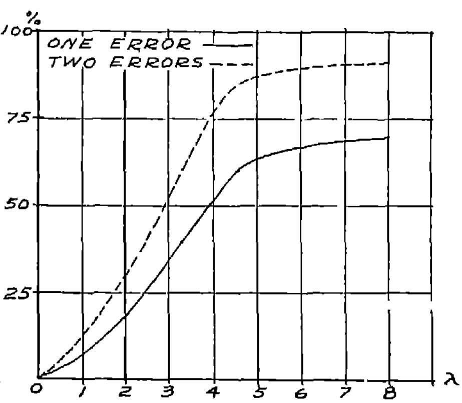


FIG 6 Comparison of the performance B_1 for one and two scalar errors in samples of size 15, $\alpha = 5\%$

5. Performance of criteria (no external estimate of σ). Criteria D_1 and D_2 have strong intuitive reasons for their use since the dispersion is estimated by s^2. The r ratios are attractive because of their simplicity and their preoccupation with the extreme values. Test F is the "studentized" ratio corresponding to B_1, and is equivalent to D_1 since $D_1 = 1 - F^2/(n - 1)$. There is no apparent difference in the performance of D_1 and r_{10} when one outlier is present and no apparent difference in D_2 and r_{20} when two outliers are present This is true for both models of contamination and for the three levels of significance investigated. However the comparison of D_2 and r_{20} was made only for $n = 5$ since critical values are not available[2] for D_2 for $n = 15$. (Critical values are available for $n \leq 12$.)

The performance of D_1 and r_{10} under the two models of contamination can be obtained by reference to the curve for r_{10} in Figure 1 and Figure 5. The curve for D_1 is practically identical with the curve for r_{10}.

[2] After this paper was submitted, the critical values of D_2 have been extended to $n \leq 20$ (see references)

There is no question that r_{10} is simpler to use, so that if this condition of contamination (scalar errors) exists, r_{10} would probably be chosen. However, as before, we should investigate what happens when more than one error is present. D_2 is designed for this case as is r_{20}. Since the performance of these two criteria is approximately the same, r_{20} would probably be chosen because of its simplicity. Critical values for this statistic are available for $n \leq 30$.

r_{11}, r_{12}, r_{20}, r_{21}, r_{22} were designed for use in situations where additional outliers may occur and we wish to minimize the effect of these outliers on the investigation of the particular value being tested.

It has been suggested that D_1 could be used repeatedly to remove more than one outlier from a sample. This procedure cannot be recommended since the presence of additional outliers handicaps the performance of both D_1 and r_{10} for small sample sizes and therefore the process of rejection might never get started. For larger sample sizes the performance of D_1 is affected much less by the presence of two errors than is the performance of r_{10}. The repetitive use of D_1 is not recommended in this case either since r_{20} performs in a superior manner to D_1 in such situations. This difference in performance of D_1 and r_{10} depends markedly on the level of significance used as well as the sample size. For small samples there is little difference in performance for any of the levels of significance one might use. For the larger sample sizes there is no appreciable difference for very high levels of significance. The difference is however very great for lower levels of significance. In fact as λ increases for two errors of the location type, the level of significance which divides the region of approach to zero performance from the region of approach to perfect performance of D_1 is given by the level of significance corresponding to a significance value of $\frac{1}{2}\left(\frac{n}{n-1}\right)$ for D_1. Thus, for example, in samples of size 15, $\frac{1}{2}\left(\frac{n}{n-1}\right) = \frac{15}{28} = .536$. This value lies between the values for the 2.5% and 5% level of significance. These values are .503 and .556 respectively. Therefore the use of the 1% or 2.5% levels will give poorer and poorer performance as λ increases, and the use of the 5% or 10% levels will give better and better performance as λ increases when two errors are present. The dividing point is such that for samples of size 11 or less the use of any of the given levels of significance will cause the performance to decrease as λ increases. For samples of size $n \leq 14$ the 1%, 2.5% and 5% levels have the same effect, and for samples of size $n \leq 16$ the 1% and 2.5%, for samples of size $n \leq 19$ just the 1% level. For three such errors the limit approached by D_1 as λ increases is $\frac{2}{3}\left(\frac{n}{n-1}\right)$. Therefore, the performance of D_1 will approach zero for all levels of significance and for all sample sizes for which critical values are known except the 10% level of significance for sample sizes larger than 21. An indication of these limiting values $\frac{k-1}{k}\,\frac{n}{n-1}$ for k contaminations present can be obtained by considering these k values to

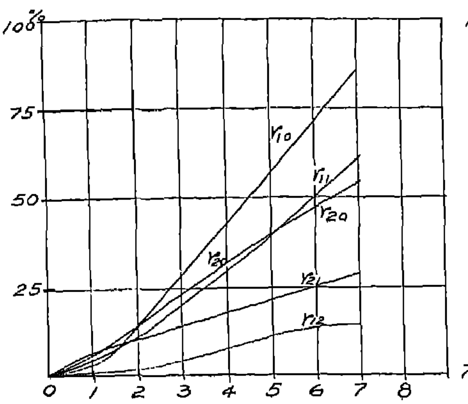


FIG. 7. Comparison of the performance of the r criteria for one location error in samples of size 5, $\alpha = 5\%$

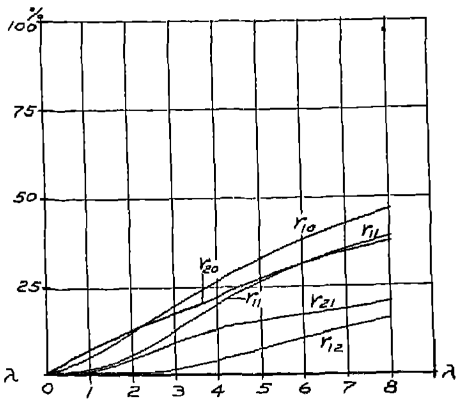


FIG. 8 Comparison of the performance of the r criteria for one scalar error in samples of size 5, $\alpha = 5\%$

be at a distance k from the population mean, computing D_1 and allowing λ to increase indefinitely.

The comparative performance of the r criteria, $\alpha = 5\%$, in samples of size 5 for the two models of contamination (one contaminator present) are given in Figures 7 and 8. For samples of size 15 the curves are given in Figures 9 and 10. A single curve suffices here since there is no discernable difference in the curves for the different r criteria. There is considerable difference in the performance curves if more than one outlier is present. However, the performances of r_{10}, r_{11}, r_{12} are essentially the same when two location outliers are present as are the performances of r_{20}, r_{21}, r_{22}. Figures 11 and 12 show the comparative performance of r_{10}, r_{11}, r_{12} for one and two contaminators for $\alpha = 5\%$ and $n = 5$. Figures 13 and 14 are for $n = 15$. Figures 15 and 16 show the comparative per-

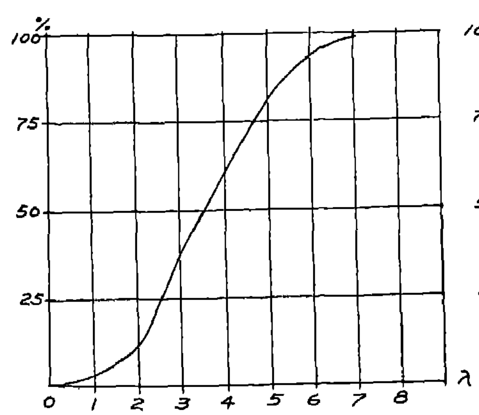


FIG. 9 Performance of the r criteria for one location error in samples of size 15, $\alpha = 5\%$

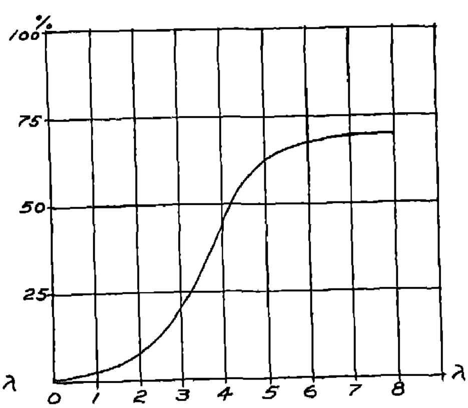


FIG. 10. Performance of the r criteria for one scalar error in samples of size 15, $\alpha = 5\%$

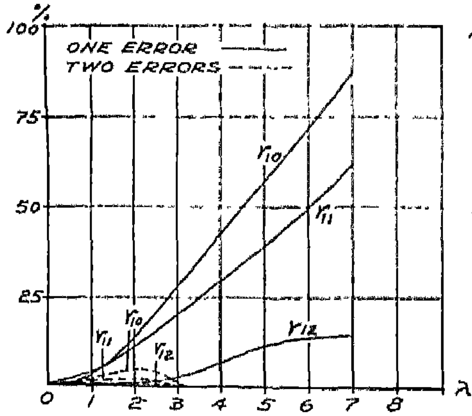


FIG. 11. Comparison of the performance of the r_1. criteria for one and two location errors in samples of size 5, $\alpha = 5\%$.

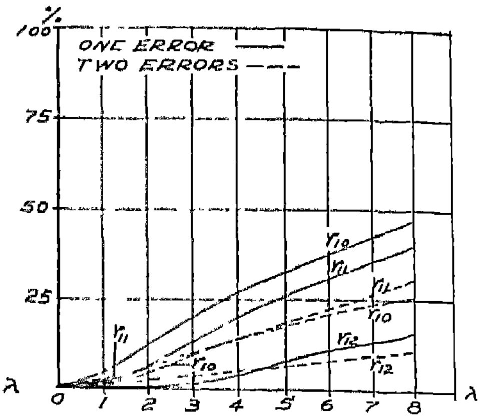


FIG. 12. Comparison of the performance of the r_1. criteria for one and two scalar errors in samples of size 5, $\alpha = 5\%$.

formance for r_{20}, r_{21}, (r_{22} is not a test for $n = 5$) for one and two contaminators for $\alpha = 5\%$ and $n = 5$. Figures 17 and 18 are for r_{20}, r_{21}, r_{22} for $n = 15$. The six curves represented by the single curve of Figure 17 lie within 5% of the curve shown. The same is true of the three curves represented by each of the two curves of Figure 18.

Since no loss in performance results for larger samples from the use of r_{20}, r_{21}, r_{22} in place of r_{10}, r_{11}, r_{12}, and further, these criteria are not appreciably affected by the presence of another outlier it would seem unwise to recommend the use of r_{10}, r_{11}, r_{12}. However, note that for small samples (see Figures 11 and 12) the performances of r_{10} and r_{11} and r_{12} are considerably better when a single

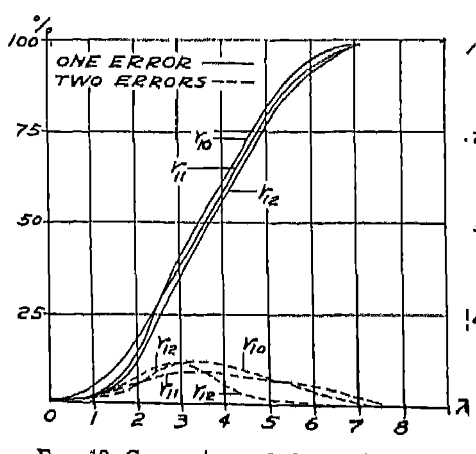


FIG. 13 Comparison of the performance of the r_1. criteria for one and two location errors in samples of size 15, $\alpha = 5\%$.

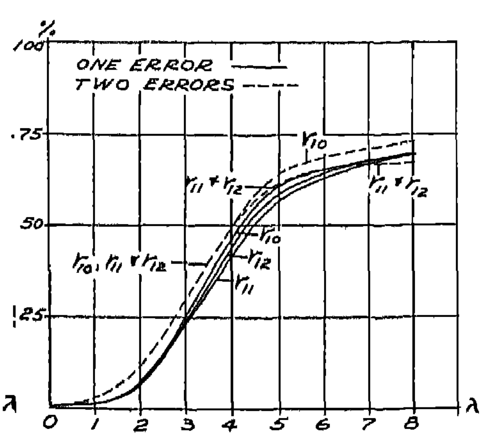


FIG. 14. Comparison of the performance of the r_1. criteria for one and two scalar errors in samples of size 15, $\alpha = 5\%$.

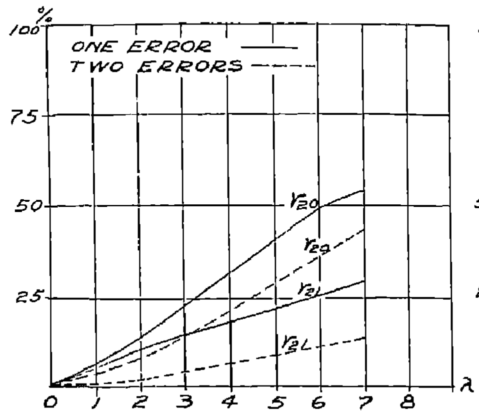


FIG 15 Comparison of the performance of the r_2 criteria for one and two location errors in samples of size 5, $\alpha = 5\%$

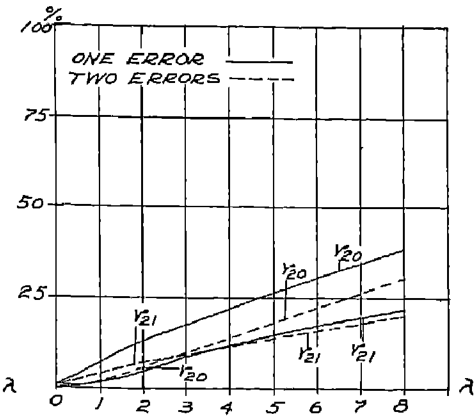


FIG 16 Comparison of the performance of the r_2 criteria for one and two scalar errors in samples of size 5, $\alpha = 5\%$

outlier is present Therefore in larger ($n > 10$) samples r_{20} or r_{21} would appear to be the best criteria In samples of size 10 or less, r_{10} or r_{20} should be used; r_{21} if the extreme value at the opposite end should be avoided.

It should be noted in the comparisons that no model of contamination was investigated which would cause one or more errors at both extremes in the sample It is obvious that the performance of D_1 and D_2 would be considerably decreased while the performance of r_{11}, r_{12}, and r_{21}, r_{22} would not be materially affected since these criteria avoid values at the opposite extreme Then repeated use might discover most of such outliers, while D_1 or D_2 might fail on the first trial.

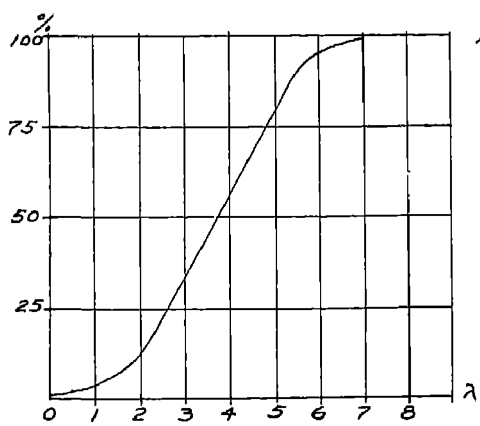


FIG 17. Comparison of the performance of the r_2 criteria for one and two location errors in samples of size 15, $\alpha = 5\%$

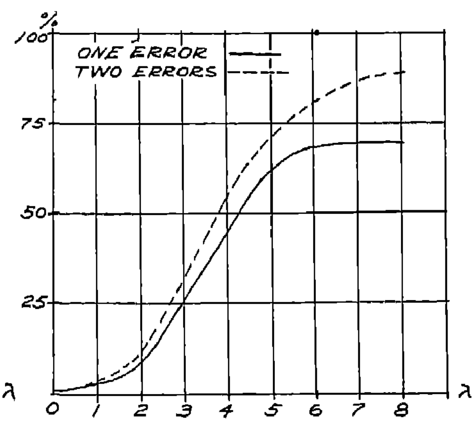


FIG. 18. Comparison of the performance of the r_2. criteria for one and two scalar errors in samples of size 15, $\alpha = 5\%$

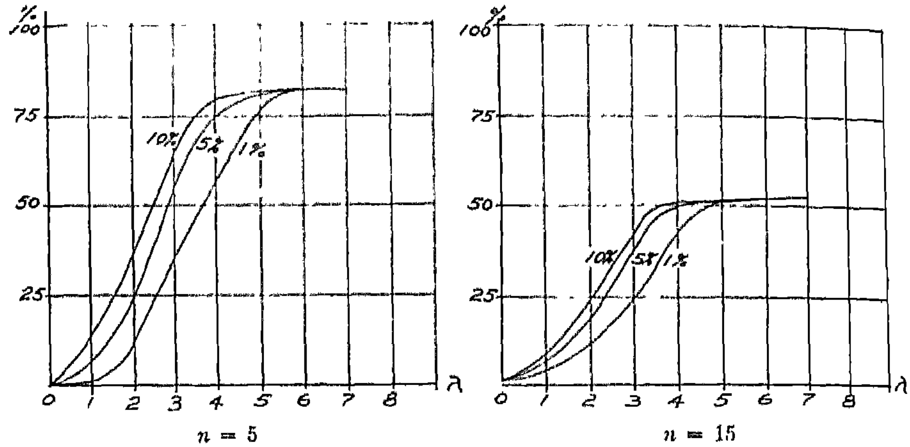


FIG. 19 Performance of B_1 for various levels of significance when the population is 10% contaminated with location errors

6. Sampling from a contaminated population. In the previous sections the performance of the various criteria were assessed for samples where a certain number of contaminators were present. One might well ask why a test is needed is it is known that contaminators are present. It would seem more realistic to state that a certain per cent of contamination will occur in the long run and that one will not know in any particular case whether 0, 1, 2, $\cdots$ contaminators will be present. One would then wish a criterion to indicate the presence of contamination in a particular sample.

The performances of these criteria will be investigated for the same two models of contamination and their performances will be reported as per cent of

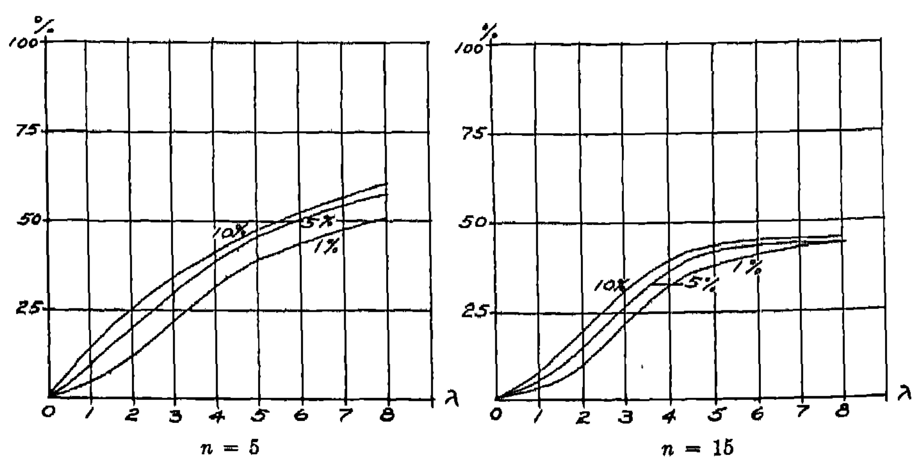


FIG. 20. Performance of B_1 for various levels of significance when the population is 10% contaminated with scalar errors

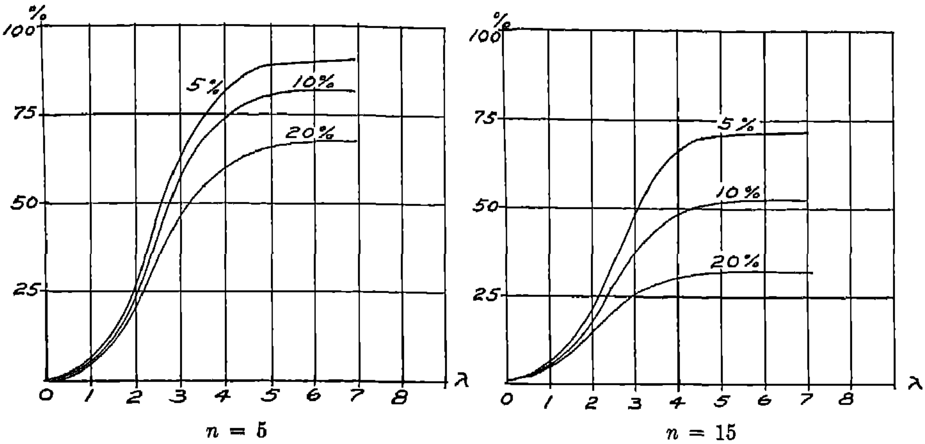


FIG. 21 Performance of B_1 for various levels of contamination for location errors and using the 5% level of significance

total contamination discovered. The tests will be applied only once to each sample. Repeated use of the criterion would in many cases increase the per cent of total contamination discovered It is not known what effect such a procedure would have on the level of significance.

Investigation has been made for 5, 10, and 20% contamination. For example, in samples of size 5 which have 10% contamination, on the average, 59.0% of the samples will contain no "errors", 32 8% will contain one, 7.3% two, 0.8% three, 0.1% four, and 0.0% five. Thus in 100 samples of 5 which are 10% contaminated with location errors having mean $\mu + 5\sigma$, about 59 contain no errors. If the r_{10} criteria is used with a 5% level of significance one value will be "dis-

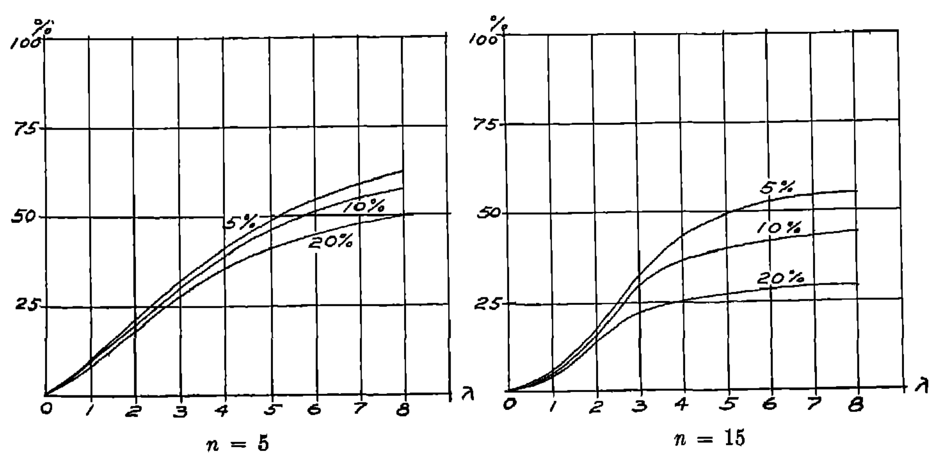
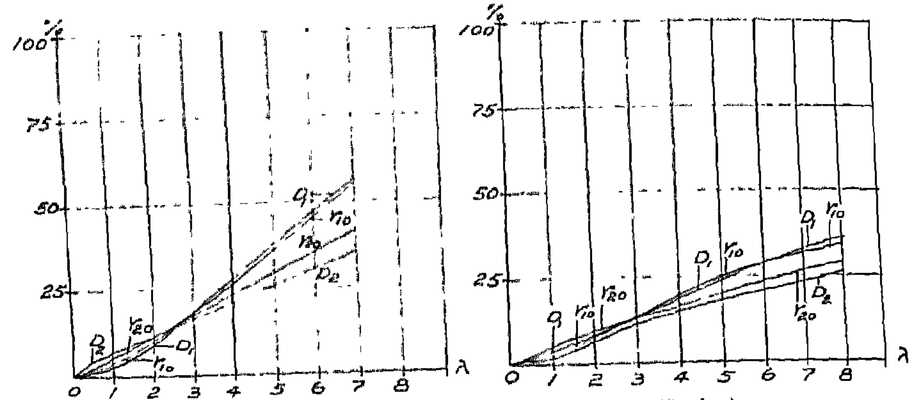


FIG. 22 Performance of B_1 for various levels of contamination for scalar errors and using the 5% level of significance.

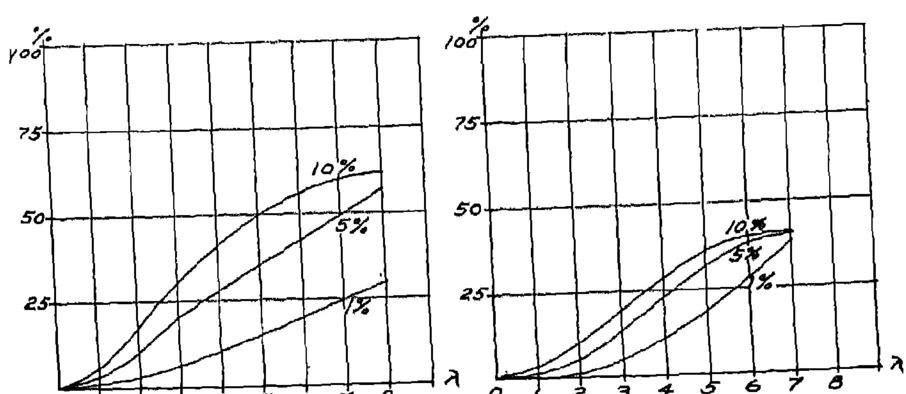


(Location) (Scalar)

FIG. 23. Performance of r_{10}, D_1, r_{20}, D_2 in samples of size 5 using the 5% level of significance and sampling from a population which is 10% contaminated.

covered" in 3.0 of the samples containing no errors. Of the 33 samples containing one "error" the "error" would by discovered in 18 of these samples. This criteria would discover none of the "errors" in samples containing more than one "error". We would have obtained 18 of the 50 contaminating values and 3 which were members of the original population.

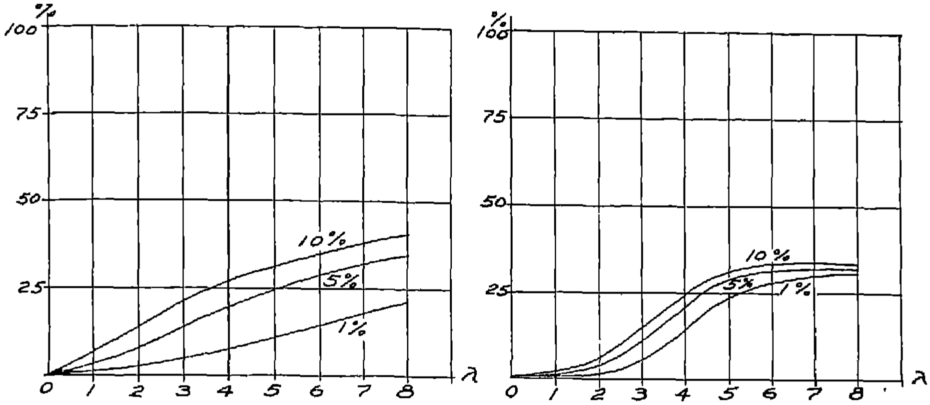
When σ is known the performance will increase when more contaminators are present. Performance however has been measured in terms of finding a single contaminator; i.e., the test has been used only once. Therefore even with increasing percent contamination the level of performance will decrease with increasing contamination. Repeated use of the test criteria has not been investigated.



$r_{10}(D_1)$
$n = 5$

$r_{22}(D_1, r_{20}, r_{21})$
$n = 15$

FIG. 24 Performance of $r_{10}(D_1)$ and $r_{22}(D_1, r_{20}, r_{21})$ for various levels of significance when the population is 10% contaminated with location errors



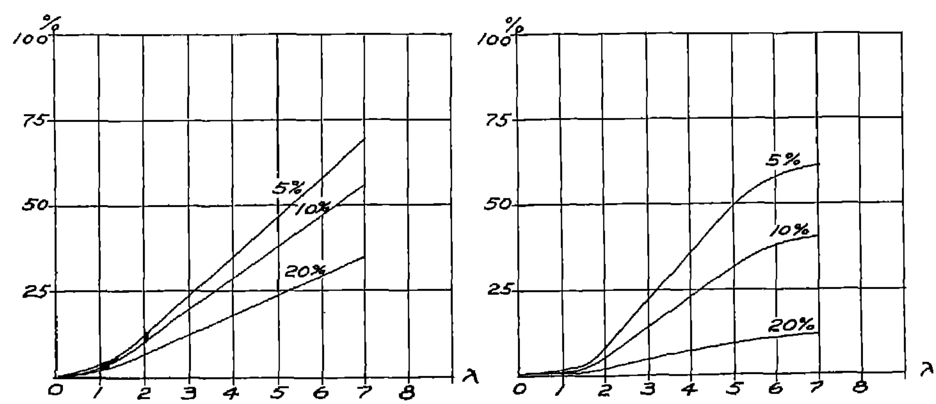
$r_{10}(D_1)$
$n = 5$

$r_{22}(D_1, r_{20}, r_{21})$
$n = 15$

FIG. 25 Performance of $r_{10}(D_1)$ and $r_{22}(D_1, r_{20}, r_{21})$ for various levels of significance when the population is 10% contaminated with scalar errors.

Criteria B_1 gives the best performance for both location and scalar errors for the levels of contamination and levels of significance considered. A and C_1 are only slightly inferior. B_2 is handicapped when more than one error is present thus its performance is poorer for heavier contamination Figure 19 shows the performance of B_1 for the different levels of significance, 10% contamination, and the two sample sizes 5 and 15 for location errors Figure 20 shows the results for scalar errors. Figures 21 and 22 show the performance of B_1 for the 5% level of significance for the different levels of contamination

When σ is not known the performance of various criteria will eventually decrease as more and more contaminators are present in the sample even though



$r_{10}(D_1)$
$n = 5$

$r_{22}(D_1, r_{20}, r_{21})$
$n = 15$

FIG 26. Performance of $r_{10}(D_1)$ and $r_{22}(D_1, r_{20}, r_{21})$ for various levels of contamination for location errors and using the 5% level of significance

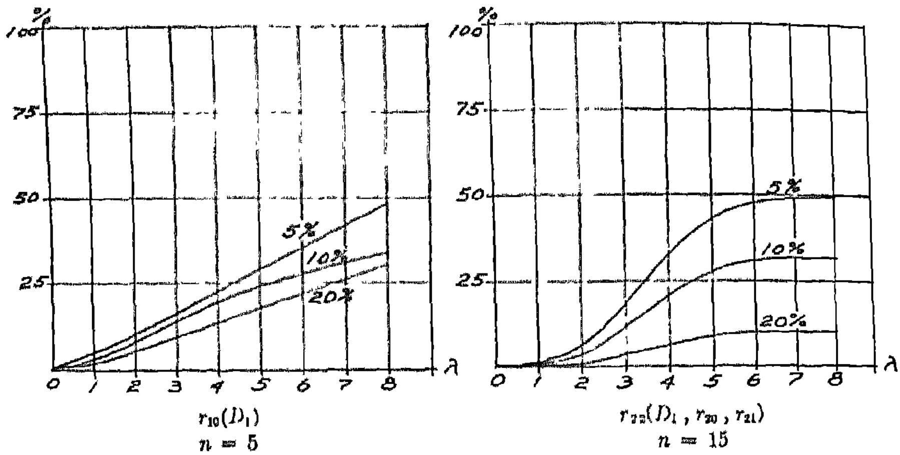


FIG. 27. Performance of $r_{10}(D_1)$ and $r_{22}(D_1, r_{20}, r_{21})$ for various levels of contamination for scalar errors and the 5% level of significance, $\alpha = 5\%$.

several of the criteria show improvement in discovering a single error if two are present. The performance of these criteria is greatly affected by the size of the sample. For samples of size 5, r_{10} and D_1 perform alike, r_{10} being superior to the other r's (r_{20} second best) for the levels of contamination considered, and D_2 is inferior to r_{20}. Figure 23 compares the performance of r_{10}, D_1, r_{20}, and D_2 for the 5% level of significance and 10% contamination. The results for other levels of significance and contamination are comparable.

For samples of size 15, r_{20}, r_{21} and r_{22} perform alike as do r_{10}, r_{11} and r_{12}. D_1 and r_{20}, r_{21}, r_{22} perform approximately the same and are superior to r_{10}, r_{11},

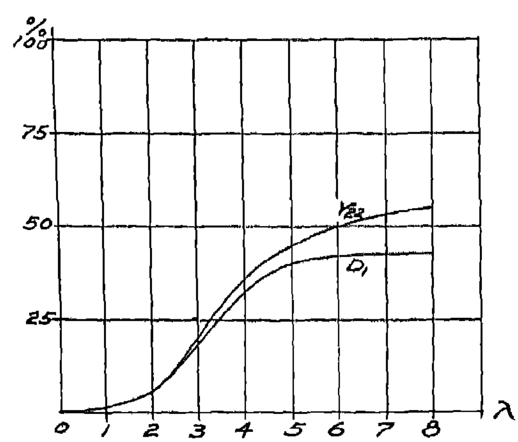


FIG 28. A comparison of the performance of r_{22} and D_1 for two scalar contaminators when tests are made at one extreme only, $\alpha = 5\%$, $n = 15$.

and r_{12}. Critical values are not available for D_2 for $n > 12$. The performances of D_1, r_{20}, r_{21} and r_{22} are indicated by a single line in Figures 24, 25, 26, and 27 which show the effect of level of significance and level of contamination of the performance of D_1, r_{20}, r_{21} and r_{22} for samples of size 15 and for r_{10} (D_1) for samples of size 5.

7. Remarks and conclusions. Throughout the investigation of performance, location errors were placed only at one extreme and scalar errors at either extreme. The test for an error was made using as a suspected value the extreme value in the direction of the location error or in the case of the scalar error the value most distant from the mean. It can be expected then that if performance were assessed when location errors could occur in either direction, different results would be obtained Also in the case of scalar errors if errors were always sought at one particular extreme or at both extremes different results would be obtained. If these changes were made in the models of contamination, those criteria designed to avoid errors at the other extreme would have an advantage over those which were not so designed for σ unknown. If σ is known the criteria which do not avoid the other extreme would have an advantage over those which do avoid the other extreme. These points just mentioned will be used to discriminate between those criteria which were judged to be equal in performance under the models used in the sampling study. For example, Figure 28 compares the performance of r_{22} and D_1 for two scalar contaminators when tests are made only at one extreme, $\alpha = 5\%$, $n = 15$.

1. For σ known:

B_1 or C_1 should be used, or in small samples A, B_1 or C_1 should be used

2. For σ unknown:

r_{10} should be used for very small samples. r_{22} should be used for sample sizes over 15. Probably r_{21} would be best for sample sizes from about 8 to 13 If simplicity in computation is not important and "errors" are not expected at both extremes D_1 would do equally well. When critical values are available for larger n, D_2 should prove useful in the larger sample sizes

LITERATURE REFERRING TO CRITERIA LISTED IN SECTION 3

(B_1) A T McKay, "The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe," *Biometrika*, Vol 27 (1935), pp. 466–471. Procedures for obtaining percentage values given.

(B_2) J. O Irwin, "On a criterion for the rejection of outlying observations," *Biometrika*, Vol 17 (1925), pp. 238–250 $Pr(B_2 > \lambda), \lambda = .1(.1)5\,0$, $n = 2, 3, 10(10)100(100)1{,}000$. Tables concerning the second and third ordered observations are also given.

(C_1) E S Pearson and H O Hartley, "The probability integral of the range in samples of n observations from the normal population," *Biometrika*, Vol 32 (1942), pp 301–310 0 1%, 0.5%, 1 0%, 2.5%, 5%, 10%, $n = 2(1)12$, values to 20 available by interpolation.

(C_2) D Newman, "The distribution of ranges in samples from a normal population, expressed in terms of an independent estimate of the standard deviation," *Biometrika*, Vol 31 (1940), pp 20–30. 1% and 5% points for C_2; for w, $n = 2(1)12, 20$; s, d.f $= 5(1)20, 24, 30, 40, 60, \infty$.

(C_2) E. S. PEARSON AND H. O. HARTLEY, "Tables of the probability integral of the studentized range," *Biometrika*, Vol. 33 (1942), pp. 89-99. Upper and lower 5% and 1% points for C_2, for w, $n = 2(1)20$; for s, d.f. $= 10(1)20, 24, 30, 40, 60, 120, \infty$.

(C_2, B_1) K. R. NAIR, "The distribution of the extreme deviate from the sample mean and its studentized forms," *Biometrika*, Vol. 35 (1948), pp. 118-144. B_1 upper and lower .1%, .5%, 1%, 2.5%, 5%, 10% points for $n = 3(1)9$.

(D_1, D_2, F, B_1) F. E. GRUBBS, "Sample criterion for testing outlying observations," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 27-58. F, D_1: 1%, 2.5%, 5%, 10%, $n \leq 25$; D_2: 1%, 2.5%, 5%, 10%, $n \leq 20$; B_1: 1%, 2.5%, 5%, 10%, $n \leq 25$.

(F) W. R. THOMPSON, "On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation," *Annals of Math. Stat.*, Vol. 6 (1935), pp 214-219. 20%, 10%, 5%, $n = 3(1)22(10)42$, 102, 202, 502, 1002.

(F) E. S. PEARSON AND CHANDRA SEKAR give a further discussion of F in "The efficiency of statistical tools and a criterion for the rejection of outlying observations," *Biometrika*, Vol. 28 (1936), pp. 308-320. 10%, 5%, 2.5%, 1%, $n = 3(1)19$.

(r's) W. J. DIXON, "Ratios involving extreme values," *Annals of Math. Stat*, to be published. r_{10}, r_{11}, r_{12}, r_{20}, r_{21}, r_{22}; .5%, 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, $n \leq 30$.

DISTRIBUTIONS RELATED TO COMPARISON OF TWO MEANS AND TWO REGRESSION COEFFICIENTS

By Uttam Chand[1]

University of North Carolina

Summary. We consider here the relative merits of different statistics available for testing two means or two regression coefficients in relation to one-sided (asymmetric) and two-sided (symmetric) alternatives in case of unequal population variances In so far as the Behrens-Fisher statistic is concerned we confine ourselves to the consideration of the behavior of its probability of Type I error in repeated sampling from populations with a fixed value of the unknown ratio of variances. In connection with the tests between two means, the present study takes its point of departure from the existing tests and investigates the question of utilizing an approximately determinate knowledge about the unknown ratio of variances In connection with the comparison of two regression coefficients and also of two linear regression functions, we consider the effect of two concomitant sources of variation, viz, the unknown ratio of residual variances and the ratio of the sums of squares of the fixed variates, on the probability of Type I and Type II errors of certain well known statistics.

1. Introduction. Consider two independent samples $x_1 \cdots x_{n_1+1}$ and $x'_1 \cdots x'_{n_2+1}$ drawn from two normal populations with means m_1 and m_2, variances σ_1^2 and σ_2^2. Let $K = \sigma_1^2/\sigma_2^2$. If K is known and $m_1 = m_2$, the quantity

$$t_K = \frac{\bar{x} - \bar{x}'}{\left[\dfrac{S(x-\bar{x})^2 + KS'(x'-\bar{x}')^2}{n_1+n_2}\left(\dfrac{1}{n_1+1} + \dfrac{1}{K(n_2+1)}\right)\right]^{\frac{1}{2}}}$$

(t_1 is Fisher's t) is distributed according to "Student's" distribution with $n_1 + n_2$ d.o.f.[2] and for the "Student's" hypothesis $H_0: m_1 = m_2$ provides a uniformly most powerful test against an asymmetric alternative $H_1: m_1 >$ (or $<$) m_2 and a type B_1 test against a symmetric alternative $H_2: m_1 \neq m_2$. If K is unknown certain approximate and exact tests have been suggested from time to time to meet this situation.

Welch [1], [2] using an approximation to the distribution of t_1 was the first to point out that if K is unknown and we assume it to be equal to unity, then the probability of Type I error of the t_1-test is subject to large variations as K varies from 0 to ∞. He also pointed out that the statistic

$$v = (\bar{x} - \bar{x}')\left[\frac{S(x-\bar{x})^2}{n_1(n_1+1)} + \frac{S'(x'-\bar{x}')^2}{n_2(n_2+1)}\right]^{-\frac{1}{2}}$$

[1] Now Assistant Professor of Mathematical Statistics at Boston University.

[2] Degrees of freedom.

which does not have "Student's" distribution for $K = 1$, has the advantage that its probability of Type I error is subject to less variation with respect to K. His approximate results were later confirmed by Hsu [3] who obtained the distribution of quantities $u_1(=t_1^2)$ and $u_2(=v^2)$ and also showed that these tests are unbiased in the sense of Neyman and Pearson. Hsu concluded on the basis of his investigations that when the sample sizes are equal and not very small, we may safely use $u_1(=u_2)$ as if K were unity. This also had been pointed out by Welch.

If on the basis of past experience some approximate value k of K were available, one would like to know if such a choice in some rough neighborhood of K would in any way improve the claim of $t_k(=t_K$ for $K = k)$ for the hypothesis $m_1 = m_2$. The distribution of this generic quantity $t_k\left(=t_1 \text{ for } k = 1; = v \text{ for } k = \frac{n_1(n_1+1)}{n_2(n_2+1)}\right)$ will be obtained in Section 2.1. It will be shown that variation in the probability of Type I error of t_k with respect to K for any k except when $t_k = v$, is essentially similar in character to that of t_1^2 [3] and is very sensitive in a neighborhood of K in which one would very often be interested (Section 2.4). This is also true of the behavior of the power function of t_k with respect to K. Consequently a t_k type of statistic will be unsuitable in general for utilizing an approximately determinate knowledge of K.

It is not possible to infer directly from Hsu's work on the relative merits of t_1 and v in relation to asymmetric aspects of "Student's" hypothesis. His basic conclusions as regards unbiasedness and the nature of variations in Type I error in the symmetric case also hold for the asymmetric case except that the Type I variations in t_1 and v are less for asymmetric than for symmetric comparisons (Section 2.5 and Table II). Furthermore it appears (Section 2.5 and Table III) that with respect to the variations of K both the asymmetric and symmetric power functions of t_1 are likely to be more sensitive than those of v. Since for equal d.o.f. both the asymmetric probability of Type I error and power function are insensitive to the vagaries of the 'nuisance' parameter K, there is an a fortiori reason for using $v(=t_1)$ as if K were unity.

Scheffé [4] considered the statistic

$$S = (\bar{x} - \bar{x}')\left(\sum_{i=1}^{n_1+1} \frac{(u_i - \bar{u})^2}{n_1(n_1+1)}\right)^{-\frac{1}{2}} \qquad (n_1 \leq n_2),$$

(equivalent to paired difference t when $n_1 = n_2$) where $u_i = x_i - \left(\frac{n_1+1}{n_2+1}\right)^{\frac{1}{2}} x'_i$ and where it is assumed that the variates in each sample have been randomized. This is essentially a "Student's" t comparison based on n_1 d.o.f. and as shown by Scheffé it is impossible to get a suitable statistic with the t-distribution with more than n_1 d.o.f. The statistic v has the t-distribution only when $K = \infty$ (n_1 d.o.f.), $K = 0$(n_2 d.o.f.) and $K = \frac{n_1(n_1+1)}{n_2(n_2+1)}$ ($n_1 + n_2$ d.o.f.). For any given n_1, n_2, K and P we can solve $P = P(v \geq t_0 \mid H_0)$ for t_0 and thus indirectly obtain

from the tabulated values of the t-distribution the number of 'effective' d.o.f. which will thus adjust v to any preassigned level of significance. We try to show in Section 2.6 that in situations where some approximate knowledge of K is available, the statistic v seems to have a decided advantage over any other statistic having the t-distribution. We show by actual computations that Welch's formula [2] provides a conservative estimate for the effective d o.f. in the light of which this comparison will be considered.

The Behrens-Fisher fiducial test employing the statistic d [5], [6], which has essentially the same structural form as v, has given rise to much controversy essentially because of inconsistencies arising from tests of significance based on the fiducial distribution of unknown parameters. We attempt to show in Section 2 7 that the fiducial test in general is 'conservative' in detecting significant results in repeated sampling from populations with a fixed value of the unknown ratio of variances.

In the case of comparison of two regression coefficients when the residual variances are unequal, we are faced with a similar type of problem. Consider two samples $y_\mu \mid x_\mu$ and $y'_\nu \mid x'_\nu$ ($\mu = 1, \cdots, n_1 + 1, \nu = 1, \cdots, n_2 + 1$), where x_μ and x'_ν are fixed and y_μ and y'_ν are normally and independently distributed according to $N(\alpha_1 + \beta_1(x_\mu - \bar{x}), \sigma_1^2)$ and $N(\alpha_2 + \beta_2(x'_\nu - \bar{x}'), \sigma_2^2)$ respectively. For the hypothesis $\beta_1 = \beta_2$ when the alternatives do not specify anything except $\beta_1 > \beta_2$ or $<\beta_2$, or $\beta_1 \neq \beta_2$ we shall consider the merits of statistics t^* and v^* which correspond to statistics t_1 and v for the two means. While the statistic t^* is sensitive to the variation of both $K = \sigma_1^2/\sigma_2^2$ and w, the ratio of the sums of squares of the fixed variates, the statistic v^* is insensitive to the variation of both. Barankin[3] has extended Scheffé's test to the comparison of two regression coefficients under the above assumptions. The statistic proposed by Barankin has Student's distribution with $n_1 - 1$ d o f. ($n_1 \leq n_2$) and provides the only exact unbiased test so far known While Scheffé's test for the comparison of two means and Barankin's test for the comparison of two regression coefficients should not be used when K is known and were never intended to utilize any available approximate information about K, the question of investigating into the possibility of using v^* in the latter situation is not without interest (Section 3). In Section 4 we consider the hypothesis of equality of two linear regression functions viz., H_0: $\alpha_1 = \alpha_2, \beta_1 = \beta_2$ when the alternatives do not specify anything except $\alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$.

In studying the behavior of the power function and the probability of Type I error of certain statistics under discussion we have made full use of Hsu's method and consequently only essential details have been given here

2. Hypothesis of equality of two means when variances are unequal

2.1. *The distribution of t_k for any values of n_1 and n_2*. Consider the test function $t_k(=t_K$ for $K = k$; Section 1) where k is some inexact value of K This can be

[3] E W. Barankin, "Extension of the Romanovsky-Bartlett-Scheffe test" *Proc. Berkeley Symposium on Math Stat. and Prob*, University of California Press, 1949, pp 433–449.

put in the form of $t_k = (\xi + \delta)(b\chi_1^2 + c\chi_2^2)^{-\frac{1}{2}}$ where ξ is $N(0, 1)$ and the χ^2's have independent χ^2-distribution with n_1 and n_2 d.o f., and where

$$\begin{aligned}
\delta &= (m_1 - m_2)\left(\frac{\sigma_1^2}{n_1+1} + \frac{\sigma_2^2}{n_2+1}\right)^{-\frac{1}{2}},\\
b &= (K/k)(n_1+n_2)^{-1}[k(n_2+1)+n_1+1][K(n_2+1)+n_1+1]^{-1},\\
c &= (n_1+n_2)^{-1}[k(n_2+1)+n_1+1][K(n_2+1)+n_1+1]^{-1},\\
b/c &= K/k.
\end{aligned}$$

In what follows we shall omit the subscript k from t_k. The joint probability element of ξ, χ_1^2 and χ_2^2 is given by

$$dF(\xi_1, \chi_1^2, \chi_2^2) = \tfrac{1}{4}(2\pi)^{-\frac{1}{2}}[\Gamma(n_1/2)\Gamma(n_2/2)]^{-1}e^{-\frac{1}{2}(\xi_1^2+\chi_1^2+\chi_2^2)}(\chi_1^2/2)^{n_1/2-1} (\chi_2^2/2)^{n_2/2-1}\,d\xi d(\chi_1^2)\,d(\chi_2^2).$$

We transform to new variables t, r and θ by the relations

$$\begin{aligned}
\xi + \delta &= t(b\chi_1^2 + c\chi_2^2)^{\frac{1}{2}},\\
b\chi_1^2 &= r^2\cos^2\theta \qquad (0 \le \theta \le \pi/2),\\
c\chi_2^2 &= r^2\sin^2\theta \qquad (-\infty \le r \le +\infty),
\end{aligned}$$

and integrate out r. To integrate out θ we put $z = \sin^2\theta$ if $b < c$ and $z = \cos^2\theta$ if $b > c$. This reduces the integration w.r.t. θ to a series of hypergeometric integrals. We finally have the following form for the frequency function of t_k:

$$g(t) = \frac{e^{-\delta^2/2}(b/c)^{n_2+1/2}c^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n_1+n_2}{2}\right)}\sum_{r=0}^{\infty}\frac{(\delta t)^r(2b)^{r/2}\Gamma\left(\frac{n_1+n_2+r+1}{2}\right)}{h!(1+bt^2)\,\frac{n_1+n_2+r+1}{2}} \cdot F\left(\frac{n_1+n_2+r+1}{2}, \frac{n_2}{2}, \frac{n_1+n_2}{2}, \frac{1-b/c}{1+bt^2}\right), \tag{2.1.1}$$

where F denotes the hypergeometric function. As a check if we put $b = c = (n_1+n_2)^{-1}$, we get the frequency function of non-central t for $n_1 + n_2$ d.o.f. For the case $b > c$ we have only to interchange b with c and n_1 with n_2.

The null distribution of $t_k(\delta = 0)$ is an even function of t_k, consequently the forms of the single and two-equal-tailed probability of Type I error will be the same except for the constant $\frac{1}{2}$. If we let $\beta_1(\delta, K, k, n_1, n_2) = \int_{t_0}^{\infty} g(t)\,dt$ denote the single upper tail power function of t_k, from (2.1.1) we obtain

$$\beta_1(\delta, K, k, n_1, n_2) = \tfrac{1}{2}e^{-\delta^2/2}(K/k)^{n_2/2}\sum_{h=0}^{\infty}\sum_{r=0}^{\infty} \cdot\frac{(\delta^2/2)^{r/2}\Gamma\left(\frac{n_2}{2}+h\right)\left(1-\frac{K}{k}\right)^h}{\Gamma\left(\frac{n_2}{2}\right)h!\,\frac{r}{2}!}\,I_{x_0}\left(\frac{n_1+n_2}{2}+h, \frac{r+1}{2}\right), \tag{2.1.2}$$

where $x_0 = (1 + bt^2)^{-1}$ and $I_{x_0}(p, q)$ is the incomplete beta ratio. To obtain the two equal tailed power function $\beta_2(\delta, K, k, n_1, n_2)$ we need only change r into $2r$ and omit the factor $\frac{1}{2}$.

2.2. *Distribution of t_k for even values of n_1 and n_2*. (For notation refer to Section 2.1). When n_1 and n_2 are even, the method of characteristic functions yields a single infinite series for the distribution of t_k, and when $\delta = 0$ this series reduces to $\frac{n_1 + n_2}{2}$ terms. The characteristic function of $X = b\chi_1^2 + c\chi_2^2$ is given by $\phi(\tau) = (1 - 2bi\tau)^{-n_1/2}(1 - 2ci\tau)^{-n_2/2}$. To obtain the form of the frequency function of X we make use of the inversion theorem and integrate round a standard contour in the lower half of the complex plane The distribution of t_k can then be obtained from the joint probability element of ξ and X. We obtain the following form for the single tailed power function of t_k:

$$\begin{aligned}\beta_1(\delta, K, k, n_1, n_2) = {} & \tfrac{1}{2}e^{-\delta^2/2}\sum_{r=0}^{\infty}\frac{(\delta^2/2)^{r/2}}{\frac{r}{2}!}\Bigg[\left(\frac{K}{K-k}\right)^{n_2/2}\sum_{h=0}^{(n_1/2)-1} \\ & \frac{(-1)^h\,\Gamma\left(\frac{n_2}{2}+h\right)}{\Gamma\left(\frac{n_2}{2}\right)h!}\left(\frac{k}{K-k}\right)^h I_{x_0}\left(\frac{n_1}{2}-h, \frac{r+1}{2}\right) \\ & + (-1)^{n_1/2}\left(\frac{k}{K-k}\right)^{n_1/2}\sum_{h=0}^{(n_2/2)-1}\frac{\Gamma\left(\frac{n_1}{2}+h\right)}{\Gamma\left(\frac{n_1}{2}\right)h!} \\ & \left(\frac{K}{K-k}\right)^h I_{x_0'}\left(\frac{n_2}{2}-h, \frac{r+1}{2}\right)\Bigg] \quad (K \geq k)\end{aligned} \tag{2.2.1}$$

where x_0 has been defined in the previous section and $x_0' = (1 + ct_0^2)^{-1}$.

2.3. *Unbiasedness of a test based on t_k*. Since the single and two tailed forms of the power function of t_k (Section 2.1) are essentially the same functions of the standardised 'distance' δ, following Hsu [3] we can show that $\frac{\partial\beta_1}{\partial\delta} \geq 0$ and $\frac{\partial\beta_2}{\partial\delta} \geq 0$ for any fixed K and k; and consequently such a generic type of statistic provides an unbiased test both against symmetric and asymmetric alternatives.

2.4. *Variations in the power function and the probability of Type I error of t_k*. For the case $k = 1$, Hsu [3] has already shown that the probability of Type I error of the statistic t_1^2 is subject to large variations w.r.t. K. He also pointed out that the behavior of the derivative of its power function w r.t. K for fixed δ was similar to that of its probability of Type I error w.r.t. K. We shall presently see that t_k also shares this property with t_1^2.

In the first place one would like to know if any choice of k in a small neighborhood of K would stabilize the variations in the Type I error of t_k to such an extent as to make it approximately insensitive to that difference between k and

K. With this end in view we shall examine the nature of variations in the probability of Type I error of t_k w.r.t K for any fixed k.

From (2.1.2) by putting $\delta = 0$ we obtain

$$P = P(t_k \geq t_0) = \tfrac{1}{2}(K/k)^{n_2/2} \sum_{h=0}^{\infty} \Gamma\left(\frac{n_2}{2} + h\right)(1 - K/k)^h \cdot \left(\Gamma\left(\frac{n_2}{2}\right)\Gamma(h+1)\right)^{-1} I_{x_0}\left(\frac{n_1+n_2}{2} + h, \tfrac{1}{2}\right). \tag{2.4.1}$$

We now differentiate (2.4.1) and after simplification obtain

$$\frac{dP}{dK} < C_1(K/k)^{-1}[n_2(n_2+1) - n_1(n_1+1)/k][K(n_2+1) + n_1 + 1]^{-1} \; (K < k).$$

Similarly

$$\frac{dP}{dK} > C_2[n_2(n_2+1) - n_1(n_1+1)/k][K(n_2+1) + n_1 + 1]^{-1} \qquad (K > k),$$

where C_1 and C_2 are certain positive constants independent of K and k.

If $\quad k = \dfrac{n_1(n_1+1)}{n_2(n_2+1)} \quad$ we have

$$\frac{dP}{dK} \lesseqgtr 0$$

for $K \lesseqgtr k$.

This is the case when t_k is identical with the statistic v defined in Section 1 and the probability of Type I error curve expressing P as a function of K has a minimum at this point: for $n_1 < n_2$ the minimum occurs for a value of $K < 1$ and vice versa. And since v is known to be insensitive to the variation of K [3], therefore t_k is insensitive to the variation of K for this value of k.

For any other assumed value of k the curve either starts decreasing from $K = \infty$ or from $K = 0$ to the point where $K = k$ depending upon the values of n_1 and n_2. In each case the ordinate of the curve continues to decrease for some distance; it may decrease to a minimum and then start increasing or else decrease indefinitely. For fixed δ the power function of t_k also has a minimum when $K = k = \dfrac{n_1(n_1+1)}{n_2(n_2+1)}$; and for any other k the behavior of its power function is similar to that of its probability of Type I error. For the case $k = 1$ numerical values of the single and two-tailed values of the probability of Type I error and power function for different values of n_1 and n_2 and K are given in Tables II and III (Section 2.5).

In certain practical situations it may happen for example that on the basis of past experience one can determine k so that $\frac{1}{2} \leq | k - K | \leq 2$. The question arises: how much is t_k sensitive to such a neighborhood for any k, K, n_1 and n_2? That it is hard to provide a practically useful answer to this question will be

apparent from the nature of the distribution of t_k, which depends both on K and k and not merely on their ratio The following Table I will indicate how in such a small neighborhood $P(t_k \geq t_0)$ can be in serious error in two different directions.

2.5. *Statistics t_1 and v in relation to asymmetric and symmetric aspects of "Student's" hypothesis.* Statistics t_1 and v are special cases of t_k and the behavior of their probability of Type I error and power function has already been discussed (Sections 2.3 and 2.4). In this section we compare the single-tailed and two tailed values of the probability of Type I error and power function in the light of several particular examples. In all these calculations e.g. in $P(t \geq t_0)$ and

TABLE I

Variations in $P(t_k \geq t_0)$ with respect to k for fixed K

($K = 5$; $n_1 = 2$, $n_2 = 4$, $t_0 = 2\,447$)

$k =$	1	2	3	4	5	6	7
	.1129	.0936	.0749	0607	.05	.0418	.0355

TABLE II

Variations in the symmetric and asymmetric probability of Type I error of v and t_1 in relation to the unknown ratio of variances K

	K	0	.125	5	1	2	4	8	16	∞	% point of tabulated t_1
	$n_1 = n_2 = 3$	.074	0633	.0504	.05	0504	0568	0633	0691	.074	single tailed 5%
$v = t$	"	092	0681	0525	.05	0525	.0597	0681	.0770	092	two-tailed 5%
	"	.034	.0181	0110	01	0110	0138	.0181	0227	034	two-tailed 1%
	$n_1 = 4$, $n_2 = 16$	.0112	0120†	.0142	0195	0227	0265	0293	0305	0324	single tailed 1%
v	"	012	.0161†	0197	.0238	.0294	0350	.0407	0433	.0465	two-tailed 1%
	$n_1 = 8$, $n_2 = 4$	.075	0087	0598	.0543	0541	.0511‡	0521	0531	056	single tailed 5%
	$n_1 = 4$, $n_2 = 16$	00011	00043	00310	01	0221	.0483	0793	0864	133	single tailed 1%
t	"	00007	.00031	.00244	01	.0310	0592	.1169	1544	222	two-tailed 1%
	$n_1 = 8$, $n_2 = 4$	.1342	.1056	0710	05	0368	.0287	.0246	.0224	.0204	single tailed 5%

† $P =$ 01 when $K =$,074
‡ $P =$ 05 when $K =$ 3 0

$P(|t| \geq t_0')$, t_0 refers to the single and t_0' to the two tailed values of Fisher's t for the appropriate number of d.o.f. Tables II and III give the approximate values for the probability of Type I error and the power function respectively both against symmetric and asymmetric alternatives.

For equal sample sizes ($v = t_1$) the Type I error and power function curves, representing probability of Type I error and power function as a function of K, have a minimum when K is unity and a maximum occurs when K is either zero or infinity. Maximum values of the probability of Type I error for several equal sample sizes are given in Table IV. It appears that for equal sample sizes the probability of Type I error and the power function are likely to be insensitive to the variation of K. We also notice in this connection that while the single

tailed values of the probability of Type I error are less than those of the two tailed values, the values of the two tailed power function for $\delta = 1$ are less than the corresponding single tailed values. This appears to be true also for the statistic v when $n_1 \neq n_2$. For unequal sample sizes also the probability of Type I error and the power function of t_1 are likely to be more sensitive to the variation of K than those of v. It may be pointed out in the sequel that while it is recognized that for unequal d.o.f. a fair comparison of the probability of Type I error and the power function of v with those of t_1 ought to adjust v and t_1 to the same level of significance, namely the same maximum (for all K) probability of Type I error, this would not alter our conclusions about the sensitive nature of t_1.

TABLE III[4]

Variations in the asymmetric and symmetric power function of t_1 and v corresponding to the 5% point of tabulated $t_1(\delta = 1)$

$K =$	0	5	1	2	∞	
$n_1 = n_2 = 3$	189	.141	.137	141	.189	symmetric
$v = t_1$	269	.229	225[5]	.229	.269	asymmetric
$n_1 = 8, n_2 = 4$	354	.202	.152	.112	.063	symmetric
t_1	428	.294	.242[6]	.194	.122	asymmetric
$n_1 = 8, n_2 = 4$	208	.196	.162	.156†	.168	symmetric
v	.286	259	.247	.244‡	.255	asymmetric

† minimum of .152 is reached for $K = 3.6$.
‡ minimum of .242 is reached for $K = 3.6$.

TABLE IV

Maximum probability of Type I error of $v(= t_1)$ for equal degrees of freedom

	Symmetric		Asymmetric	
$n_1 + 1 = n_2 + 1$	5%	1%	5%	1%
7	.0721	.0224	.0625	.0182
9	0668	.0193	.0595	.0162
11	.0635	0173	.0576	.0150
15	.0598	.0152	.0555	.0136
21	.0569	.0137	.0538	.0125

2.6. *Statistic v, Scheffé's test and paired difference t.* If K is known, v or Scheffé's statistic S should not be used. If K is unknown, S is an ingenious device for getting a Student's t with $\min(n_1, n_2)$ d.o.f. and provides the only exact unbiased test so far known. In such a situation since nothing is known about K, a fair comparison of the power function of S with v ought to adjust v to the same maximum probability of Type I error for all K (maximum will occur for $K = 0$ or $K = \infty$ according as $n_1 \gtrless n_2$); and at such a maximum significance level it is

[4] The author acknowledges with pleasure the help given in the preparation of this table by Miss Elizabeth Shuhany of the Statistical Laboratory, Boston University

[5] Values taken from [7]

recognized that v cannot be uniformly better than S For samples of equal size n the use of the paired difference t with $n - 1$ d.o.f (equivalent to S when $n_1 = n_2$; Section 1) provides a suitable test for two reasons: (i) it is exact and (ii) as shown by Walsh [8] has a high power efficiency.

If any approximate a priori information about K is available, v appears to be the only suitable statistic to utilize such information While S was not intended to cope with such a situation, t_k (Section 2 4) has been shown to be unsuitable. Since v is insensitive to the variation of K, we shall not be far wrong in using 'effective' d.o.f. based upon an assumed value k of K satisfying some such relation as $\frac{1}{2} \leq | k - K | \leq 2$. The effective d o.f. of v as given by Welch [1] and as given by $P = P(v \geq t_0)$ or by $P = P(| v | \geq t_0')$ for fixed P (listed in Table V as calculated d.o f) are identical for $K = 0, 1$, an ∞ $(n_1 = n_2)$ and (ii) $K = 0, \dfrac{n_1(n_1 + 1)}{n_2(n_2 + 1)}$, and ∞ $(n_1 \neq n_2)$. For other values of K it appears from Table V that Welch's formula errs on the conservative side. The effective number of d o f. vary between $n_1 + n_2$ and $\min(n_1, n_2)$ (cf. d o.f. for S). Consequently in the absence of any

TABLE V

Adjusted power function of v in the light of 'effective' degrees of freedom

Sample Size	Adjusted asymmetric power function of v for probability of Type I error of .05								Effective d.o f.							
	$\delta = 1$				$\delta = 2$				Calculated				Welch's formula			
	$K = 0$	.125	4	∞	$K = 0$	125	4	∞	$K = 0$	125	4	∞	$K = 0$	.125	4	∞
$n_1 + 1 = n_2 + 1 = 3$	174	204	204	174	384	476	.476	.384	2	3 30	3 36	2	2	2 94	2 94	2
$n_1 + 1 = n_2 + 1 = 7$	225	.236	236	225	550	.581	581	550	6	9 14	9 14	6	6	8 82	8 82	6
$n_1 + 1 = 9, n_2 + 1 = 5$	210	227	242	233	504	.556	504	572	4	6 50	11 90	8	4	5 14	11 90	8

best unbiased test and in the light of any approximate information about K it would appear that v has a decided advantage over any other statistic.

2.7. *The Behrens-Fisher test in repeated sampling.* Consider the statistic

$$d = (\bar{x} - \bar{x}')\,(s_1^2 + s_2^2)^{-\frac{1}{2}} = t_1 \sin\theta - t_2 \cos\theta,$$

where s_1^2 and s_2^2 are the unbiased estimates of the variances of the means $\bar{x}$ and $\bar{x}'$ respectively, t_1 and t_2 have independent "Student's" distributions with n_1 and n_2 d.o.f. respectively, and $\tan\theta = s_1/s_2$. On the basis of the "fiducial" distribution of σ_1^2 and σ_2^2 Fisher [6] regards d as a "mixture" of t_1 and t_2 with constant coefficients. It is to be noted that if s_1 and s_2 are fixed in the classical sense t_1 and t_2 have independent normal conditional distributions with zero means and variances σ_1^2/s_1^2 and σ_2^2/s_2^2 respectively; and if s_1 and s_2 vary in their own distribution d is identical with v (Section 1).

Neyman [9] considered the integral of the joint probability law of $\bar{x}$, $\bar{x}'$, s_1^2, s_2^2 over the set $\dfrac{| \bar{x} - \bar{x}' |}{\sqrt{s_1^2 + s_2^2}} \leq t_1 \sin\theta - t_2 \cos\theta$ where the quantity on the right also depends upon s_1 and s_2 and is the quantity d tabulated by Sukhatme [10], [11]

Neyman showed in particular that if pairs of normal populations with different K are sampled ($n_1 + 1 = 13$, $n_2 + 1 = 7$), then the relative frequency of correct statements about $m_1 - m_2$ based on the 5% points of d will not be equal to the expected .95 and will vary with K.

We consider here the following similar type of question: what is the nature of discrepancies that will arise in the probability of Type I error by the repeated use of the Behrens-Fisher test in sampling from two normal populations? We observe that since d and v have the same structural form, the appropriate probability of Type I error in such a situation will be given by the probability integral of v (Sections 2.2 and 2.5).

TABLE VI

Minimum and maximum† values of $P(\mid v \mid \geq d_0)$ for different values of K

K	0	.05	.1	.2	∞	d_0
$n_1 + 1 = n_2 + 1 = 7$	.05	.0321	.0307	.0321	.05	2.447
	.0508	**.0329**	**.0313**	**.0329**	**.0508**	2.435
$n_1 + 1 = n_2 + 1 = 9$	.05	.0362	.0346	.0362	.05	2.306
	.0512	**.0367**	**.0358**	**.0367**	**.0512**	2.292
$n_1 + 1 = n_2 + 1 = 13$	.05	.0405	.0396	.0405	.05	2.179
	.0507	**.0434**	**.0403**	**.0434**	**.0507**	2.170
$n_1 + 1 = 7,\ n_2 + 1 = 13$	.0307	.0281	.0317	.0393	.05	2.447
	.05	**.0460**	**.0516**	**.0597**	**.0720**	2.179
$n_1 = n_2 = \infty$	.05	.05	.05	.05	.05	1.960

† maximum values have been indicated in bold type

We observe that $P(\mid v \mid \geq x)$ is a monotone decreasing function of x for any fixed K, n_1 and n_2. Furthermore for fixed x, n_1 and n_2 we have $\frac{dP}{dK} \gtreqless 0$ for (i) $K \gtreqless 1$, $n_1 = n_2$ and (ii) $K \gtreqless \frac{n_1(n_1+1)}{n_2(n_2+1)}$, $n_1 \neq n_2$. Table VI gives the minimum and maximum values of $P(\mid v \mid \geq d_0)$ for different values of K where d_0 corresponds to the highest and lowest value of tabulated d. It appears that for equal sample sizes the minimum probability of Type I error is less than .05 and will converge to .05 when K is either infinity or zero. The maximum probability of Type I error converges to a value slightly higher than .05. This probability also converges to .05 with increasing size of equal samples for every K. For unequal sample sizes e.g. $n_1 < n_2$, the minimum values converge to .05 when $K = \infty$ and if $n_1 > n_2$, this convergence takes place when $K = 0$. The maximum values are both greater and less than .05.

3. Hypothesis of equality of regression coefficients when residual variances are unequal.

3.1. *Unbiasedness of tests based on statistics t^* and v^*.* Consider

$$t^* = (b_1 - b_2)\left[\frac{S(y - Y)^2 + S'(y' - Y')^2}{n_1 + n_2 - 2}\left(\frac{1}{M_1} + \frac{1}{M_2}\right)\right]^{-\frac{1}{2}}$$

and

$$v^* = (b_1 - b_2)\left[\frac{S(y - Y)^2}{M_1(n_1 - 1)} + \frac{S'(y' - Y')^2}{M_2(n_2 - 1)}\right]^{-\frac{1}{2}},$$

where b_1 and b_2 are regression coefficients calculated from independent samples; Y and Y' are the sample regression functions; $M_1 = S(x - \bar{x})^2$ and $M_2 = S'(x' - \bar{x}')^2$. Under the assumptions of Section 1 these two quantities are distributed as

$$t^* = (\xi + \Delta)\ (\mu_1\chi^2_{1,n_1-1} + \mu_2\chi^2_{2,n_2-1})^{-\frac{1}{2}},$$
$$v^* = (\xi + \Delta)\ (\lambda_1\chi^2_{1,n_1-1} + \lambda_2\chi^2_{2,n_2-1})^{-\frac{1}{2}},$$

respectively, where ξ is $N(0, 1)$ and the χ^2's have independent χ^2-distribution with d.o.f. indicated in the second subscripts, and where

$$\begin{aligned} M_1/M_2 &= w, \\ \mu_1 &= K(w + 1)\ (K + w)^{-1}\ (n_1 + n_2 - 2)^{-1}, \\ \mu_2 &= (w + 1)\ (K + w)^{-1}\ (n_1 + n_2 - 2)^{-1}, \\ \frac{\mu_1}{\mu_2} &= K, \\ \Delta &= (\beta_1 - \beta_2)\left(\frac{\sigma_1^2}{M_1} + \frac{\sigma_2^2}{M_2}\right)^{-1}, \\ \lambda_1 &= K(K + w)^{-1}\ (n_1 - 1)^{-1}, \\ \lambda_2 &= w(K + w)^{-1}\ (n_2 - 1)^{-1}, \\ \frac{\lambda_1}{\lambda_2} &= (K/w)\ \frac{n_2 - 1}{n_1 - 1}. \end{aligned}$$

Consequently these two statistics have the same basic distribution as obtained previously for t_k (Section 2.1) and their power functions are monotone increasing functions of the standardized 'distance' Δ for fixed values of K, w, n_1 and n_2. While the statistic t^* has "Student's" distribution with $n_1 + n_2 - 2$ d.o.f. whenever $K = 1$, the statistic v^* is only so distributed when $K = w(n_1 - 1)(n_2 - 1)^{-1}$.

3.2. *Variations in the probability of Type I error and power function of t^* and v^*.* The behavior of the partial derivatives of the probability of Type I error and the power function of t^* and v^* w.r.t. K and also in relation to w is essentially the same. For purposes of illustration we shall only consider the behavior of the probability of Type I error. We shall presently see that for the hypothesis $\beta_1 = \beta_2$ (cf. "Student's" hypothesis $m_1 = m_2$) while t^* is sensitive to the variation of K and w, v^* is insensitive to both.

3.2.1. *Variations w.r.t. K for fixed w* Remembering that the χ^2's in the denominator of t^* have respectively $n_1 - 1$ and $n_2 - 1$ d.o.f., we can write down $P(t^* \geq t_0)$ from the corresponding form for t_1 (Section 2.3). After simplification we obtain

$$\frac{\partial P}{\partial K} < L_1[(n_2 - 1) - w(n_1 - 1)]\ (K + w)^{-1}/K \qquad (K < 1), \tag{3.2.1.1}$$

where $z_0 = (1 + \mu_1 t_0^2)^{-1}$. If we make use of the relation $P(n_1, n_2, M_1, M_2, K) = P(n_2, n_1, M_2, M_1, K^{-1})$ in (3.2.1.1) we obtain

$$\text{(3.2.1.2)} \qquad \frac{\partial P}{\partial K} > L_2(K + w)^{-1}[(n_2 - 1) - w(n_1 - 1)] \qquad (K > 1),$$

where L_1 and L_2 are certain positive constants independent of M_1, M_2 and K.

Similarly for the statistic v^* we have

$$\text{(3.2.1.3)} \qquad \frac{\partial P}{\partial K} < D_1(K\phi)^{-1}[(n_2 - 1) - w(n_1 - 1)\phi]/(K + w) \qquad (K\phi < 1)$$

and

$$\text{(3.2.1.4)} \qquad \frac{\partial P}{\partial K} > D_2[(n_2 - 1) - w(n_1 - 1)\phi]/(K + w) \qquad (K\phi > 1),$$

where D_1 and D_2 are certain positive constants independent of K, M_1 and M_2 and where $\phi = \dfrac{n_2 - 1}{w(n_1 - 1)}$. We notice that if (i) $n_1 = n_2$ and $w = 1$ or (ii) $w = \dfrac{n_2 - 1}{n_1 - 1}$, we have $t^* = v^*$ and both from (3.2.1.1), (3.2.1.2) and from (3.2.1.3), (3.2.1.4) we obtain $\dfrac{\partial P}{\partial K} \lesseqgtr 0$ for $K \lesseqgtr 1$. In the case (i) the maximum probability of Type I error occurs at $K = \infty$ and $K = 0$. In case (ii) the maximum will sometimes occur for $K = 0$ and sometimes for $K = \infty$, depending on the relative magnitude of n_1 and n_2.

For other situations t^* and v^* exhibit a type of behavior essentially similar to that of t_1 and v (Section 2.5) We notice that the (P, K) curve for v^* has a minimum when $K = \dfrac{w(n_1 - 1)}{n_2 - 1}$. If $n_1 = n_2$, the minimum point is given by $K = w$ Therefore with an approximate knowledge of K, a useful practical hint to remember is to so adjust M_1 and M_2 as to have w approximately equal to K. If $n_1 \neq n_2$ any information about σ_1^2 being greater or less than σ_1^2 can be used with decided advantage to adjust M_1, M_2, n_1 and n_2 so as to reduce considerably the risk of the first kind and thus work in a region of the (P, K) curve where there is not much danger of bias in the probability of Type I error. This will also reduce the fluctuations of the power function of v about its minimum which also occurs for $K = \dfrac{w(n_1 - 1)}{n_2 - 1}$.

3.2.2. *Variations in relation to w for fixed K.* The partial derivative of $P(t^* \geq t_0)$ with respect to w is given by

$$\text{(3.2.2.1)} \qquad \begin{aligned} \frac{\partial P}{\partial w} &= \tfrac{1}{2}(1 - K)K^{n_2 - 1/2}(K + w)^{-1} \sum_{h=0}^{\infty} (1 - K)^h \\ &\quad \cdot \frac{\Gamma\left(\frac{n_2 - 1}{2} + h\right) z_0^{(n_1 + n_2 - 2)/2 + h} (1 - z_0)^{\frac{1}{2}}}{h!\,\Gamma\left(\frac{n_2 - 1}{2}\right) B\left(\frac{n_1 + n_2 - 2}{2} + h, \frac{1}{2}\right)} \qquad (K < 1). \end{aligned}$$

Therefore

$$\frac{\partial P}{\partial w} > 0$$

for $K < 1$.
Similarly

$$\frac{\partial P}{\partial w} \leq 0$$

for $K \geq 1$.
To justify the differentiation of the series in (3.2 2.1) we make use of the result

$$I_{z_0}\left(\frac{n_1 + n_2 - 2}{2} + h, \tfrac{1}{2}\right) - I_{z_0}\left(\frac{n_1 + n_2 - 2}{2} + h + 1, \tfrac{1}{2}\right) = \frac{z_0^{(n_1+n_2-2)/2+h}(1 - z_0)^{\frac{1}{2}}}{\left(\frac{n_1 + n_2 - 2}{2} + h\right) \mathrm{B}\left(\frac{n_1 + n_2 - 2}{2} + h, \tfrac{1}{2}\right)},$$

and consequently the series under consideration may be shown to be dominated by an absolutely and uniformly convergent series for $0 < K < 1$

For the statistic v^* consider

$$\begin{aligned} P(v^* \geq t_0) &= \tfrac{1}{2}(K\phi)^{(n_2-1)/2} \sum_{h=0}^{\infty} (1 - K\phi)^h \, \Gamma\left(\frac{n_2 - 1}{2} + h\right) \\ &\cdot \left[\Gamma(h + 1)\Gamma\left(\frac{n_2 - 1}{2}\right)\right]^{-1} I_{y_0}\left(\frac{n_1 + n_2 - 2}{2} + h, \tfrac{1}{2}\right) \quad (K\phi < 1) \end{aligned} \tag{3.2.2.2}$$

where $y_0 = (1 + \lambda_1 t_0^2)^{-1}$. We notice from (3.2.2.2) and from the form of quantities λ_1 and λ_2 (Section 3.1) that $P(v^* \geq t_0)$ depends on K and w only through the product of K and $1/w$. Consequently variations of P w.r.t. $1/w$ for fixed K are the same as those of P w.r t. K for fixed w. Thus we may directly infer that $P(v^* \geq t_0)$ will be insensitive to the variations of w. The following Table VII will illustrate the nature of variations in the probability of Type I error in the tests based on t^* and v^* in relation to w.

TABLE VII

Variations in the probability of Type I error of t^* and v^*

($K = 2$; $n_1 = n_2 = 7$; $t_0 = 1.782$)

w	0	.25	.5	1	2	∞
$P(t^* \geq t_0)$	0259	.0358	.0427	0512	.0594	.0866
$P(v^* \geq t_0)$	0625	0570	.0539	0512	.05	.0625

It would appear that on the analogy of statistics t_1 and v for the comparison of two means one could guess about the sensitive nature of t^* in relation to the

variations of the 'nuisance' parameter K. The additional drawback in t^* which stems from the monotone nature of its variations with respect to w is a further warning against the use of a t^* type statistic for the hypothesis $\beta_1 = \beta_2$ when $\sigma_1^2 \neq \sigma_2^2$.

4. Hypothesis of equality of two linear regression functions when variances are unequal.

4.1. *The statistic Z.* (For notation refer to Sections 2.1 and 3.1). Consider the model given in Sections 1 and 3 for the comparison of two regression coefficients. If the variances are equal, the statistic based on the likelihood ratio criterion for the composite hypothesis $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$ is given by

$$Z = \frac{(\bar{y}_1 - \bar{y}_2)^2(n_1 + 1)(n_2 + 1)(n_1 + n_2 + 2)^{-1} + (b_1 - b_2)^2 M_1 M_2 (M_1 + M_2)^{-1}}{S(y - Y)^2 + S'(y' - Y')^2}.$$

The quantity Z is distributed like the ratio of two independently distributed χ^2's and consequently its distribution is precisely determined under the hypothesis. If $\sigma_1^2 \neq \sigma_2^2$, Z can be put in the form of

$$Z = (a_1 \chi_{1,1}^2 + a_2 \chi_{2,1}^2)\ (K\chi_{3,n_1-1}^2 + \chi_{4,n_2-1}^2)^{-1},$$

which is now distributed as the ratio of 'mixtures' of independently distributed χ^2's with d.o.f. indicated in the second subscripts and where

$$a_1 = [n_1 + 1 + K(n_2 + 1)]\ (n_1 + n_2 + 2)^{-1},$$

$$a_2 = (K + w)\ (1 + w)^{-1}.$$

In the non-null case when $\alpha_1 \neq \alpha_2$, $\beta_1 \neq \beta_2$ the numerator of Z is a mixture of non-central squares. If we let $\beta(K, w, \delta, \Delta, n_1, n_2)$ denote the power function of Z, following Robbins and Pittman [12] we obtain

$$\beta(K, w, \delta, \Delta, n_1, n_2) = \sum_{j=0}^{\infty}\sum_{h=0}^{\infty}\sum_{k=0}^{\infty} c_j\, d_h p_k I_\zeta \left(\frac{n_1 + n_2}{2} + h - 1, k + j + 1\right) \tag{4.1.1}$$

$$\left(K > 1,\ w < \frac{n_1 + 1}{n_2 + 1}\right),$$

where

$$c_j = \frac{(a_1/a_2)^{\frac{1}{2}}\Gamma(j + \frac{1}{2})}{\Gamma(\frac{1}{2})j!}\ (1 - a_1/a_2)^j,$$

$$d_h = \frac{K^{-(n_1-1)/2}\Gamma\left(\frac{n_1 - 1}{2} + h\right)\left(1 - \frac{1}{K}\right)^h}{\Gamma\left(\frac{n_1 - 1}{2}\right)h!},$$

$$p_k = e^{-\frac{1}{2}D^2}(\tfrac{1}{2}D^2)^k/k! \qquad (D^2 = \delta^2 + \Delta^2),$$

$$\zeta = (1 + Z_0/a_1)^{-1}.$$

4.2. *Variations in the probability of Type I error and the power function of Z* Corresponding to (4.1.1) we obtain the expression for the probability of Type I error $P(Z \geq Z_0)$ by putting $D = 0$ and $k = 0$. It has not been possible to establish any definite law concerning the behavior of the probability of Type I error and the power function w.r.t. the 'nuisance' parameter K. However we shall presently establish their monotone dependence on the variable parameter w.

We differentiate $P(Z \geq Z_0)$ with respect to w and after simplification obtain

$$\frac{\partial P}{\partial w} = (K - 1)(a_1/a_2)^{\frac{1}{2}} \Sigma\Sigma \frac{d_h \Gamma(j + \frac{1}{2})}{j! \Gamma(\frac{1}{2})} \left[\frac{1}{2}\left(1 - \frac{a_1}{a_2}\right)^j - \frac{a_1}{a_2} j \left(1 - \frac{a_1}{a_2}\right)^{j-1} \right]$$

$$\cdot I_\xi\left(\frac{n_1 + n_2}{2} + h - 1, j + 1\right) = \frac{(K - 1)(a_1/a_2)^{\frac{1}{2}}}{(K + w)(1 + w)} \Sigma\Sigma\, d_h \frac{\Gamma(j + 3/2)}{j! \Gamma(\frac{1}{2})}$$

$$\cdot \left[I_\xi\left(\frac{n_1 + n_2}{2} + h - 1, j + 1\right) - I_\xi\left(\frac{n_1 + n_2}{2} + h - 1, j + 2\right) \right] < 0$$

for $K > 1$, $w < \dfrac{n_1 + 1}{n_2 + 1}$. Similarly by utilizing an appropriate expression for $P(Z \geq Z_0)$ for $K > 1$, $w > \dfrac{n_1 + 1}{n_2 + 1}$ we can show that $\dfrac{\partial P}{\partial w} < 0$. For the case $K < 1$ it can be shown that $P(Z \geq Z_0)$ is a monotone increasing function of w. This is also true of the dependence of the power function of Z on w.

4.3. *Unbiasedness of Z.* We differentiate (4.1.1) w.r.t. δ and Δ and after simplification obtain $\dfrac{\partial\beta}{\partial\delta} \geq 0$, $\dfrac{\partial\beta}{\partial\Delta} \geq 0$. Thus the power function of Z has a relative minimum at $\delta = 0$, $\Delta = 0$

The author is greatly indebted to Professors Harold Hotelling and William G. Madow for guidance in this research and to the referees for many useful suggestions and criticisms.

REFERENCES

[1] B. L. Welch, "The significance of the difference between two means when the population variances are unequal", *Biometrika*, Vol 29 (1938), pp 350–361.

[2] M G. Kendall, *The Advanced Theory of Statistics*, Vol 2, C. Griffin and Co., 1946, pp 112–115.

[3] P. L. Hsu, "Contribution to the theory of "Student's" *t*-test as applied to the problem of two samples", *Stat. Res Memoirs*, Vol. 2 (1938), pp 1–24

[4] H. Scheffé, "On the solutions of the Behrens-Fisher problem based on the *t*-distribution", *Annals of Math Stat.*, Vol 14 (1943), pp 35–44

[5] W. V. Behrens, "Ein Beitrag zur fehlerberechnung bei wenigen Beobachtungen", *Landw. Jb.*, Vol 68 (1929), pp 807–837.

[6] R A. Fisher, "The fiducial argument in statistical inference", *Annals of Eugenics*, Vol 6 (1935), pp. 391–398

[7] J Neyman, "Statistical problems in agricultural experiments", *Jour. Roy Stat Soc , Supp.*, Vol 2 (1935), pp 107–180

[8] J. E. WALSH, "On the power efficiency of a t-test formed by pairing sample values" *Annals of Math. Stat*, Vol 38 (1947), pp. 601-604

[9] J. NEYMAN, "Fiducial argument and the theory of confidence intervals", *Biometrika*, Vol. 32 (1941), pp. 128, ff

[10] P. V. SUKHATME, "On Fisher and Behrens' test of significance for the difference in means of two normal samples", *Sankhyā*, Vol 4 (1938), pp. 39–48

[11] R. A FISHER AND F YATES, *Statistical Tables*, Oliver and Boyd, 1943.

[12] H. ROBBINS AND E. J. G. PITMAN, "Application of the method of mixtures to quadratic forms in normal variates", *Annals of Math. Stat.*, Vol 20 (1949), pp 552–560

THE EXTREMAL QUOTIENT

By E. J. Gumbel and R. D. Keeney

New York City and Metropolitan Life Insurance Company

Summary. The extremal quotient is defined as the ratio of the largest to the absolute value of the smallest observation. Its analytical properties for symmetrical, continuous and unlimited distributions are obtained from a study of the auto-quotient defined as the ratio of two non-negative variates with identical distributions. The relation of the two statistics is established by proving that, for sufficiently large samples from an initial distribution with median zero, the largest (or smallest) value may be assumed to be positive (or negative) and that the extremes are independent. It follows that the distribution and the probability of the extremal quotient possess certain symmetries, and that its median is unity. As many moments exist for the extremal quotient as moments and reciprocal moments exist simultaneously for the initial variate The logarithm of the extremal quotient is symmetrically distributed. These properties hold for all continuous symmetrical unlimited variates which possess a monotonically increasing probability function.

For the exponential type, the asymptotic distribution of the extremal quotient can only be expressed by an integral. In this case, no moments exist. For the Cauchy type, the asymptotic distribution is very simple, and the logarithm of the extremal quotient has the same distribution as the midrange for initial distributions of the exponential type.

It is not necessary to consider asymmetrical distributions since, in this case, for sufficiently large samples, one of the extremes will outweigh the other, unless the distribution is nearly symmetrical or has rapidly varying tails.

1. The auto-quotient and the extremal quotient. Let x and y be two independent non-negative continuous variates, unlimited to the right Let $f_1(x)$ and $f_2(y)$ be the distributions (probability densities), and let $F_1(x)$ and $F_2(y)$ be the probability functions. Then the joint distribution of the two variates is their product The quotient

$$Q = x/y \tag{1.1}$$

is also non-negative and unlimited to the right. Since

$$x = yQ, \qquad \frac{dx}{dQ} = y,$$

the joint distribution $\mathrm{w}(y, Q)$ of the quotient Q and the variate y is

$$\mathrm{w}(y, Q) = f_1(yQ)f_2(y)\cdot y, \tag{1.2}$$

and the marginal distribution $h(Q)$ of the variate Q alone becomes

$$h(Q) = \int_0^\infty y f_1(yQ) f_2(y)\, dy. \tag{1.3}$$

The quotient Q possesses a mode if (and only if) $f_1(x)$ possesses a mode.

Assume now that the two variates x and y have the same distribution

$$f_1(x) = f(x); \qquad f_2(y) = f(y) \tag{1.4}$$

with the same parameter values. The quotient of two variates with identical distributions is henceforth called the *auto-quotient* q_a. It may be realized if there are two independent series of observations taken from the same population and ordered in time. Each value from the first series is divided by the corresponding value from the second series. Another realization consists in dividing each value obtained in one series of independent observations by every other value. A third realization is obtained by considering two asymmetrical distributions $f_1(x)$ and $f_2(y)$ where $x \geqq 0$, $y \leqq 0$, and

$$f_2(y) = f_1(-x). \tag{1.4'}$$

The two distributions are called mutually symmetrical, and the auto-quotient is

$$q_a = x/(-y).$$

From the definition of the auto-quotient it follows that the distribution of q_a must be the same as the distribution of its reciprocal $r = 1/q_a$. The proof of this statement is simple. Under the condition (1.4), the distribution $h(q_a)$ becomes, from (1.3)

$$h(q_a) = \int_0^\infty y f(yq_a) f(y)\, dy. \tag{1.5}$$

The distribution $h_1(r)$ of the reciprocal is

$$h_1(r) = \frac{1}{r^2} \int_0^\infty y f(y/r) f(y)\, dy.$$

If y/r is replaced by x, the distribution of r is

$$h_1(r) = h(q_a). \tag{1.6}$$

Thus, the distribution of the auto-quotient of a non-negative unlimited variate is invariant under a reciprocal transformation.

The shape of the distribution $h(q_a)$ and the location of the mode may be obtained from the density of probability $h(1/q_a)$ at the value $1/q_a$ (which differs, of course, from the distribution $h_1(r)$ of $r = 1/q_a$). From (1.5) we obtain

$$h(1/q_a) = \int_0^\infty y f(y/q_a) f(y)\, dy.$$

The transformation

$$y/q_a = z, \qquad dy = q_a\,dz,$$

leads to

$$h(1/q_a) = q_a^2 h(q_a). \tag{1.7}$$

This is a *symmetry relation* for the distribution of the auto-quotient of a non-negative unlimited variate. If q_a is larger than unity,

$$h(1/q_a) > h(q_a). \tag{1.8}$$

If the distribution $h(q_a)$ is continuous for all values of q_a, the derivative of equation (1.7) with respect to q_a leads, for $q_a = 1$, to

$$h'(1) = -h(1) \tag{1.9}$$

If the distribution $h(q_a)$ possesses a unique mode, it must be less than unity.

The moments $\overline{q_a^k}$ are, from (1.5)

$$\overline{q_a^k} = \int_{q_a=0}^{q_a=\infty} \int_{y=0}^{y=\infty} q^k y f(qy) f(y)\, dy\, dq$$
$$= \int_{y=0}^{y=\infty} \frac{f(y)}{y^k} \int_{q_a=0}^{q_a=\infty} (q_a y)^k f(q_a y)\, d(q_a y)\, dy.$$

The inner integral is the moment $\overline{y^k}$ of order k of the initial variate y, and the remaining integral is its reciprocal moment $\overline{y^{-k}}$ of order $-k$. Thus

$$\overline{q_a^k} = \overline{y^k}\;\overline{y^{-k}} = \overline{q_a^{-k}}. \tag{1.10}$$

The moments of order k and of order $-k$ of q_a exist if the moments and the reciprocal moments of order k for the initial variate exist simultaneously. The second equation in (1.10) also follows immediately from the invariance of q_a under a reciprocal transformation. Even if the initial distribution possesses all moments, the mean $\bar{q}_a$ need not exist, and the same holds, of course, for the mean error and the higher moments The procedure, usual in economic and meteorological statistics, of calculating the quotients of two series of independent positive variables in order to test whether this ratio is constant may be misleading, especially if the two series happen to be samples taken from the same population. The theoretical mean need not exist, and the calculated mean of the observed quotients need not characterize the relation between the two series.

The probability function $H(Q)$ of the quotient Q obtained from (1.3) is

$$H(Q) = \int_0^Q \int_0^\infty y f_1(zy) f_2(y)\, dy\, dz.$$

Change of the order of integration leads to

$$H(Q) = \int_0^\infty f_2(y) F_1(Qy)\, dy$$

The probability function $H(q_a)$ of the auto-quotient obtained from (1.4) is

$$H(q_a) = \int_0^1 F(q_a y)\, dF \tag{1.11}$$

Integration by parts leads to

$$H(q_a) = 1 - q_a \int_0^\infty F(y) f(q_a y)\, dy. \tag{1.12}$$

The boundary condition, $H(0) = 0$; $H(\infty) = 1$ can immediately be verified if the preceding equation is written in the form

$$H(q_a) = 1 - \int_0^\infty F(z/q_a) f(z)\, dz \tag{1.13}$$

The probability $H(q_a)$ possesses a symmetry relation which is analogous to (1.7). The probability at the value $1/q_a$ is, from (1.11),

$$H(1/q_a) = \int_0^\infty F(y/q_a) f(y)\, dy.$$

If we introduce the variable of integration

$$y = q_a z,$$

we obtain from (1.12)

$$H(q_a) = 1 - H(1/q_a). \tag{1.14}$$

If q_a is any quantile, such that $H(q_a) = P$, its reciprocal $1/q_a$ has the probability $1 - P$. The first quartile (decile) is the reciprocal of the third quartile, (ninth decile) and so on.

For $q_a = 1$, equation (1.14) leads to

$$H(1) = \tfrac{1}{2}. \tag{1.14'}$$

The median of the auto-quotient of a positive unlimited variate is unity. From (1.9) it follows that the median surpasses the mode, if a unique mode exists

Finally, equation (1.14) may be used to construct a symmetrical distribution. If a new variate

$$z = \lg q_a \tag{1.15}$$

with the probability function $H^*(z)$ is introduced, the symmetry relation (1.14) becomes

$$H^*(z) = 1 - H^*(-z). \tag{1.16}$$

The logarithm of the auto-quotient of a positive unlimited variate has a symmetrical distribution about median zero. The geometric mean of q_a exists and is equal to unity.

These results hold if each observed value of a non-negative unlimited variate is divided by each other observed value. They do not hold for the quotients of two specific order statistics because, in general, the fundamental assumption of independence does no longer hold. However, some consequences for the quotients of extreme mth values may be deduced.

Consider a symmetrical unlimited variate. Then the distribution ${}_m\varphi({}_mx)$ of the mth smallest value ${}_mx$, and the distribution $\varphi_m(x_m)$ of the mth largest value x_m are mutually symmetrical in the sense of (1.4′) Therefore the extremal quotient

$$q_m = \frac{x_m}{-{}_mx} \tag{1.17}$$

may be interpreted as an auto-quotient provided that 1) the probability for x_m to be negative, and ${}_mx$ to be positive, may be neglected; 2) the distributions of the mth smallest and the mth largest values are independent. Under these conditions the distribution, the moments, and the probability function of the extremal quotient are obtained from (1.5), (1.10), and (1.11) respectively, if the initial distribution $f(y)$ is replaced by the distribution of the mth largest values $\varphi_m(x_m)$ The symmetry relations (1.7) and (1.14) and their consequence, that the median is equal to unity, hold in particular for $m = 1$, i.e for the extremal quotient proper.

The validity of the two conditions has now to be established.

a) Consider a symmetrical distribution $f(x)$ with median zero. Then the probability that the largest among n observations, x_n, is equal to or less than a certain x, is $1 - F^n(x)$ The probability P that the largest among n values is positive, i.e. larger than the median, is

$$P = 1 - 2^{-n}. \tag{1.18}$$

If n is sufficiently large, this probability differs from unity by an amount that can be made as small as we please Even for relatively small samples, say $n = 20$, the probability that the largest value will be positive is of the order $1 - 10^{-6}$. Thus, we expect only one largest value in a million samples of size 20 to be negative The same argument shows that the smallest value x_1 may be expected to be negative. Thus the postulate

$$x_n \geqq 0; \qquad x_1 \leqq 0, \tag{1.19}$$

is a very weak restriction upon the sample size. If m is sufficiently small, the same result holds for the mth extremes.

b) It is known [7] that the joint distribution $\mathfrak{w}_n(x_1, x_n)$ of the extremes taken from an initial distribution of the exponential type converges, for sufficiently large samples, toward the product of the asymptotic distribution $\varphi(x_n)$ of the largest value, and ${}_1\varphi(x_1)$ of the smallest value. A similar theorem will now be proven for a general class of continuous distributions.

Let ${}_mx$ be the mth smallest observation; let x_l be the lth largest observation where m and l are small compared to n, n being large. Then the joint distribution $\mathfrak{w}_n({}_mx, x_l)$ is

$$\mathfrak{w}_n({}_mx, x_l) = \frac{n!}{(m-1)!(l-1)!(n-m-l)!} \tag{1.20}$$
$$F({}_mx)^{m-1}(F(x_l) - F({}_mx))^{n-m-l}(1 - F(x_l))^{l-1}f({}_mx)f(x_l)).$$

Now the transformation

$$n(1 - F(x_l)) = \xi; \qquad nF({}_mx) = \eta; \qquad 0 \leqq \xi \leqq n, \qquad 0 \leqq \eta \leqq n, \tag{1.21}$$

due to Cramér ([1], p. 371) is used. Then the joint distribution $v_n(\xi, \eta)$ of the new variates ξ and η becomes

$$v_n(\xi, \eta) = \frac{n!}{n^2(m-1)(l-1)!(n-m-l)!}\left(\frac{\xi}{n}\right)^{m-1}\left(1 - \frac{\xi - \eta}{n}\right)^{n-m-l}\left(\frac{\eta}{n}\right)^{l-1},$$

where $m + l$ is small compared to n. As n increases, $v_n(\xi, \eta)$ converges to

$$v(\xi, \eta) = \left(\frac{\xi^{m-1}e^{-\xi}}{(m-1)!}\right)\left(\frac{\eta^{l-1}e^{-\eta}}{(l-1)!}\right),$$

so that in the limit ξ and η are independent. If now the mild restriction is imposed that $F(x)$ be monotonically increasing, (1.21) defines a one to one transformation, and therefore there must exist an inverse function uniquely defining ${}_mx$ as a function of ξ, and x_l as a function of η. From the limiting independence of ξ and η the limiting independence of the extremes ${}_mx$ and x_l follows at once.

Thus the second condition is fulfilled, and the mth extremal quotient shares all properties of the auto-quotient. This holds also for initial symmetrical distributions which do not possess asymptotic distributions of the extremes.

In the following, the two types of initial distributions of an unlimited variate are considered for which asymptotic distributions of the extremes exist, namely, the exponential and the Cauchy type. For simplicity, only the extremal quotient proper, designated by q, is studied. The two asymptotic probabilities of the extremal quotients for these symmetrical distributions are obtained by introducing the asymptotic distributions of the largest value into the probability function (1.11) of the auto-quotient.

2. Application to the exponential type. For symmetrical distributions of the exponential type the asymptotic distribution of the largest value is

$$\varphi(x) = \alpha \exp\left[-\alpha(x - u) - e^{-\alpha(x-u)}\right], \tag{2.1}$$

where u and α are defined in terms of the initial probability $F(x)$ and the initial distribution $f(x)$ by

$$F(u) = 1 - 1/n; \qquad \alpha = nf(u), \tag{2.2}$$

n being the sample size. The distribution (2 1) will now be simplified by introducing a new parameter λ defined by

$$e^{\alpha u} = \lambda > 0. \tag{2.3}$$

To see the meaning of λ, consider Laplace's first distribution, then the so called logistic [6], and the normal distributions, all of which are of the exponential type In the first two cases we obtain, from (2.2), after some calculations,

$$(2.4)\quad \alpha = 1, \qquad u = \lg n - \lg 2; \qquad \alpha = 1 - 1/n, \qquad u = \lg (n - 1),$$

whereas for the normal distribution, we have asymptotically

$$\alpha = u = \sqrt{2 \lg (n/\sqrt{2\pi})}$$

and

$$(2.4')\qquad \lambda = n^2/(2\pi).$$

For these distributions, and interpreted in this sense, λ is of the order of the sample size or its square.

From (2.3) and (2.1) the distribution $\varphi(x)$ and the probability function $\Phi(x)$ are

$$(2.5)\qquad \varphi(x) = \alpha\lambda \exp [-\alpha x - \lambda e^{-\alpha x}]; \qquad \Phi(x) = \exp [-\lambda e^{-\alpha x}].$$

In order to fulfill the condition (1.19), namely $\Phi(0) = 0$, the distribution $\varphi(x)$ must be truncated at $x = 0$. This leads to the truncated distribution $\varphi_t(x)$ and the truncated probability $\Phi_t(x)$ where

$$(2.6)\quad \varphi_t(x) = \frac{\alpha\lambda \exp [-\alpha x - \lambda e^{-\alpha x}]}{1 - e^{-\lambda}}; \qquad \Phi_t(x) = \frac{\exp [-\lambda e^{-\alpha x}] - e^{-\lambda}}{1 - e^{-\lambda}}.$$

The asymptotic probability function $H_\lambda(q)$ for the extremal quotient of a symmetrical variate of the exponential type is now obtained from (1.11), if y, $f(y)$, and $F(y)$, are replaced by x, $\varphi_t(x)$ and $\Phi_t(x)$, respectively, and the index a is dropped. Consequently, from (2.6),

$$H_\lambda(q) = \frac{1}{(1 - e^{-\lambda})^2} \int_0^\infty \alpha x \exp [-\alpha x - \lambda e^{-\alpha x} - \lambda e^{-\alpha q x}]\, dx - \frac{e^{-\lambda}}{(1 - e^{-\lambda})^2} \int_0^\infty \alpha\lambda \exp [-\alpha x - \lambda e^{-\alpha x}]\, dx.$$

The transformation

$$e^{-\alpha x} = z; \qquad \alpha e^{-\alpha x}\, dx = -dz$$

leads to

$$(2.7)\qquad H_\lambda(q) = \frac{1}{(1 - e^{-\lambda})^2} \int_0^1 \lambda e^{-\lambda(z + z^q)}\, dz - \frac{e^{-\lambda}}{1 - e^{-\lambda}}.$$

This probability of the extremal quotient for initial symmetrical distributions of the exponential type is not truely asymptotic since the parameter λ depends upon n. (See Addendum).

Unfortunately, the expression (2.7) cannot be integrated Therefore the probability function has to be studied in an analytic way. For this purpose we first recall the general properties

$$H(0) = 0; \qquad H(1) = \tfrac{1}{2}; \qquad H(\infty) = 1,$$

valid for any value of λ Furthermore, for any λ, we have the symmetry relation (1.14). These properties can be verified at once from (2.7).

The numerical values of $H_\lambda(q)$ can easily be calculated for $q = \frac{1}{2}$ and $q = 2$. Consider a value of λ, say of the order 6. Then formula (2.7) may be written

$$H_\lambda(2) = \int_0^1 \lambda e^{-\lambda(z+z^2)}\,dz \tag{2.8}$$
$$= \sqrt{\lambda}\, e^{\lambda/4} \int_0^1 e^{-\lambda(z+\frac{1}{2})^2} \sqrt{\lambda}\,dz.$$

If we introduce

$$\sqrt{\lambda}\,(z + \tfrac{1}{2}) = \frac{t}{\sqrt{2}}; \qquad \sqrt{\lambda}\,dz = \frac{dt}{\sqrt{2}},$$

the probability $H_\lambda(2)$ becomes a difference of two normal probability integrals,

$$H_\lambda(2) = \sqrt{\pi\lambda}\, e^{\lambda/4}\left[1 - F\left(\sqrt{\frac{\lambda}{2}}\right) - \left(1 - F\left(3\sqrt{\frac{\lambda}{2}}\right)\right)\right],$$

where F stands for the normal probability function.

The second expression may be neglected compared to the first one for $\lambda \geqq 4$, whence

$$H_\lambda(2) = \sqrt{\frac{\lambda}{2}}\, e^{\lambda/4} \int_{\sqrt{\lambda/2}}^{\infty} e^{-t^2/2}\,dt. \tag{2.9}$$

The symmetry relation (1.14) leads to the knowledge of $H_\lambda(\frac{1}{2})$. Thus the three probabilities $H_\lambda(\frac{1}{2})$, $H(1)$, and $H_\lambda(2)$ are known.

To see the influence of λ on $H_\lambda(2)$, we use a method due to R. D. Gordon [4]. This author considers a function R_x defined by

$$R_x = e^{x^2/2}\int_x^{\infty} e^{-t^2/2}\,dt, \quad x > 0, \tag{2.10}$$

and proves that

$$\frac{dR}{dx} = xR - 1 < 0; \qquad \frac{d^2R}{dx^2} = x\frac{dR}{dx} + R > 0.$$

It follows that

$$\frac{d}{dx}(xR) > 0.$$

If we substitute $\sqrt{\lambda/2}$ for x, this inequality may be written, from (2.9) and (2.10),

$$\frac{d}{d\sqrt{\frac{\lambda}{2}}}\left(\sqrt{\frac{\lambda}{2}}\, e^{\lambda/4}\int_{\sqrt{\lambda/2}}^{\infty} e^{-t^2/2}\,dt\right) = 2\sqrt{2\lambda}\,\frac{dH_\lambda(2)}{d\lambda} > 0.$$

Consequently $H_\lambda(2)$ increases with λ whereas, from (1.14), the probability $H_\lambda(\frac{1}{2})$ decreases with λ. The following table gives the probabilities $H_\lambda(2)$ and $H_\lambda(\frac{1}{2})$; (2.9) and their differences

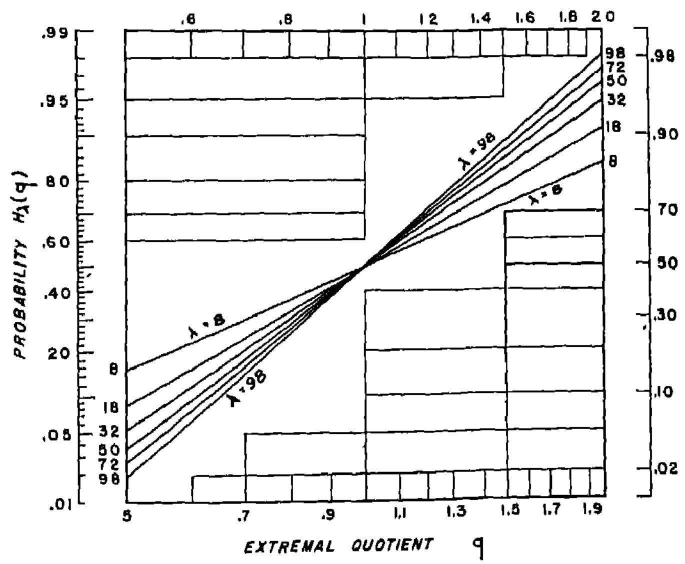
$$P_\lambda(2) = H_\lambda(2) - H_\lambda(\tfrac{1}{2}). \tag{2.11}$$

Asymptotic probabilities of the extremal quotient for symmetrical distributions of the exponential type

Parameter	Probabilities (2.9), (1.14)		Probability (2.11)
λ	$H_\lambda(2)$	$H_\lambda(\frac{1}{2})$	$P_\lambda(2)$
8	.84376	.15624	.68752
18	.91377	.08623	.82754
32	.94661	.05339	.89322
50	.96438	.03562	.92876
72	.97427	.02573	.94854
98	.98087	.01913	.96174

The approximative shape of $H_\lambda(q)$ is traced, for $\lambda = 8, \ldots, 98$, and $\frac{1}{2} < q < 2$ in Graph (1). Since we know from (1.16) that lg q has a symmetrical distribution, we use a logarithmically normal probability paper where q is plotted on the abscissa in a logarithmic scale, and $H_\lambda(q)$ is plotted on the ordinate in a normal probability scale The probability $P_\lambda(2)$ for any value of q to be contained in the interval $\frac{1}{2} < q < 2$ increases with λ, i.e., with the sample size, and the distribution of the extremal quotient contracts.

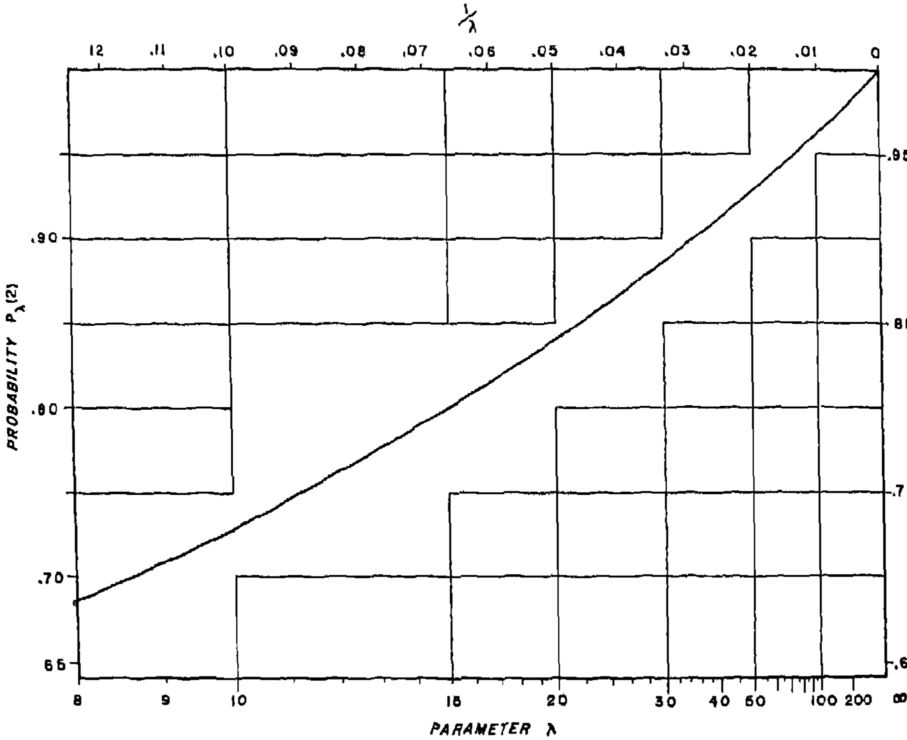
1) ASYMPTOTIC PROBABILITY OF THE EXTREMAL QUOTIENT FOR THE EXPONENTIAL TYPE



If the initial distribution is unknown, the parameter λ has to be estimated from the observed extremal quotients. Equation (2.11) may be used for this

purpose. We calculate the observed relative frequency $P_\lambda(2)$ of extremal quotients contained between $q = \frac{1}{2}$ and $q = 2$, and substitute it for the probability $P_\lambda(2)$. To facilitate this estimate of λ, we trace $P_\lambda(2)$ against λ in graph (2). The probability $P_\lambda(2)$ is traced on the ordinate in linear scale, and the parameter λ is traced on the abscissa in inverse scale. Thus λ is easily estimated from the observed relative frequency $P_\lambda(2)$.

2) ESTIMATION OF THE PARAMETER λ



The distribution $h_\lambda(q)$ of the extremal quotient obtained by differentiating the probability function (2.7) with respect to q is

$$h_\lambda(q) = \frac{1}{(1 - e^{-\lambda})^2} \int_0^1 \lambda^2 e^{-\lambda(z+z^q)} z^q(-\lg z)\, dz. \tag{2.12}$$

The symmetry relation (1.7) is easily verified. We now investigate the boundary value $h_\lambda(0)$ and prove that

$$\lim_{q \to 0} h_\lambda(q) = h_\lambda(0). \tag{2.13}$$

This is not obvious, since z^q becomes indeterminate if both z and q vanish. For the proof of (2.13), consider the integral

$$\mathrm{I} = \lambda \int_0^1 e^{-\lambda z}(-\lg z)\, dz \tag{2.14}$$

or

$$\mathrm{I} = (1 - e^{-\lambda}) \lg \lambda - \gamma + e^{-\lambda} \lg \lambda - ei(-\lambda). \tag{2.15}$$

The last term, the exponential integral, is positive. The value of $h_\lambda(0)$ is thus, from (2.12)

$$h_\lambda(0) = \frac{\lambda e^{-\lambda}(\lg \lambda - \gamma - ei(-\lambda))}{(1 - e^{-\lambda})^2}, \tag{2.16}$$

The difference

$$\Delta = (1 - e^{-\lambda})^2 (h_\lambda(q) - h_\lambda(0))$$

becomes, from (2.12), (2.15) and (2.16), by the use of the mean value theorem and after expansion

$$\Delta = f(\lambda) \int_0^1 (e^{-\lambda z^q} z^q - e^{-\lambda})\, dz$$

$$= f(\lambda) \sum_{\nu=0}^{\infty} \frac{(-1)^\nu \lambda^\nu}{\nu!} \left(\frac{1}{(\nu + 1)q + 1} - 1 \right),$$

where $f(\lambda)$ is a positive function. Since the series is absolutely convergent, the difference Δ vanishes for $q = 0$, and the density of probability for $q = 0$ is given by (2.16). The condition $h_\lambda(0) \geqq 0$, valid for any distribution, is met provided that

$$\lambda > 1.794 \tag{2.17}$$

By virtue of (2.4) this is a (weak) condition concerning the sample size. From (2.16) it follows that $h_\lambda(0)$ does not vanish although its numerical value is very small.

The existence of at least one mode follows from the fact that the distribution $h_\lambda(q)$ is continuous, very small for $q = 0$, and vanishes for $q = \infty$. Equation (1.9) proves that any mode is inferior to unity. The distribution contracts for increasing values of the parameter Therefore the mode approaches the median with increasing sample size.

Since the distributions of the exponential type do not possess reciprocal moments it follows from (1.10) that the distribution $h_\lambda(q)$ does not possess moments. The mean extremal quotient $\bar{q}$ diverges. Because the logarithmically normal distribution used in graph (1) as first approximation to the distribution $h_\lambda(q)$ possesses all moments, the distribution $h_\lambda(q)$ has a much longer tail than the logarithmically normal one.

3. Application to the Cauchy type. For the exponential type, the asymptotic distribution of the extremal quotient can only be expressed in the form of an integral containing a parameter λ which is a function of the sample size. For the Cauchy type, to be defined in the following, the asymptotic distribution will turn out to be very simple.

A distribution of a variate $x \geqq 1$ was said [5] to be of the Pareto type if

$$\lim_{x=\infty} x^k(1 - F(x)) = A; \qquad k > 0; \qquad A > 0. \tag{3.1}$$

We now say that a variate is of the *Cauchy type* if it is unlimited, continuous, subject to (3.1), and symmetrical about zero. Distributions of the Pareto and the Cauchy type do not possess moments of an order equal to or larger than k However, not all unlimited symmetrical distributions with a finite number of moments are of the Cauchy type.

The simplest example of such a distribution is the Cauchy distribution itself

$$f(x) = \frac{1}{\pi(1 + x^2)}; \qquad F(x) = \tfrac{1}{2} + \frac{1}{\pi} \text{ arc tg } x, \tag{3.2}$$

which possesses no moments. For large absolute values of x, the usual expansion leads to

$$F(x) = 1 - \frac{1}{\pi x} + O(x^{-2}); \qquad F(-x) = \frac{1}{\pi x} - O(x^{-2}).$$

If the factors $O(x^{-2})$ are neglected, the parameters A and k in (3.1) are

$$A = \pi^{-1}; \qquad k = 1. \tag{3.2'}$$

For the Cauchy type, the asymptotic probability $\Pi(x)$ and distribution $\pi(x)$ of the largest value $x = x_n$ established by Fréchet [3], R. A. Fisher [2] and R. von Mises [8] are

$$\Pi(x) = \exp\left[-\left(\frac{u}{x}\right)^k\right]; \qquad \pi(x) = \frac{k}{u}\left(\frac{u}{x}\right)^{k+1} \exp\left[-\left(\frac{u}{x}\right)^k\right], \tag{3.3}$$

where u is defined by (2.2).

The condition (1.19) is fulfilled for any sample size which is so large that the asymptotic distribution of the extremes may be used. The asymptotic probability $H_k(q)$ of the extremal quotient for the Cauchy type is obtained from (1.11), if y, $f(y)$ and $F(y)$ are replaced by x, $\pi(x)$, and $\Pi(x)$, respectively, where the indices n and a are omitted. Consequently, from (3.3),

$$H_k(q) = \int_0^\infty \frac{k}{u}\left(\frac{u}{x}\right)^{k+1} e^{-(u/x)^k - (u/qx)^k}\, dx.$$

From the transformation

$$\left(\frac{u}{x}\right)^k = z; \qquad \frac{k}{u}\left(\frac{u}{x}\right)^{k+1} dx = dz,$$

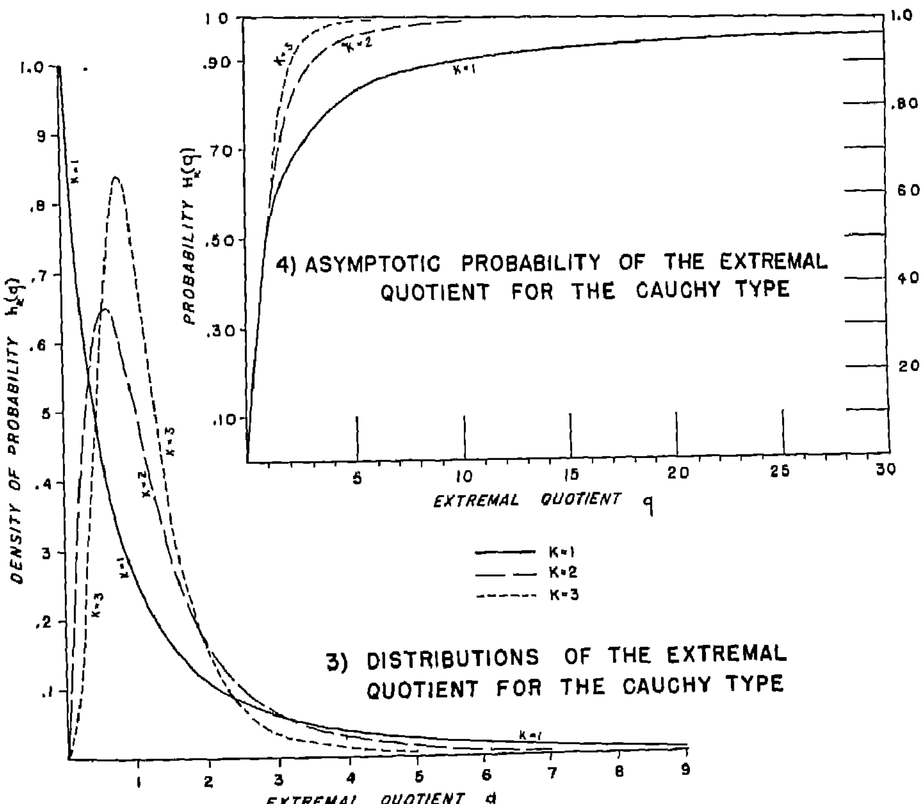
the asymptotic probability $H_k(q)$ and the asymptotic distribution $h_k(q)$ of the extremal quotient become simply

$$H_k(q) = \frac{q^k}{1 + q^k}; \qquad h_k(q) = \frac{kq^{k-1}}{(1 + q^k)^2}, \qquad q \geqq 0. \tag{3.4}$$

Evidently, the symmetry relations (1.7) and (1.14) are fulfilled for any k. The graphs (3) and (4) show the distribution $h_k(q)$ and the probability $H_k(q)$ for the most interesting cases $k = 1, 2, 3$. From

$$\frac{d \lg H_k(q)}{dk} = \lg q(1 - H_k(q))$$

it follows: For k increasing, the probability $H_k(q)$ decreases for $q < 1$, and increases for $q > 1$. *The distribution contracts with increasing values of the parameter* k as shown in the graphs (3) and (4). The more moments that exist in the initial distribution, the more concentrated is the distribution of the extremal quotient.



The density of probability

$$h_k(1) = k/4$$

of the median obtained from (3.4) and (1.14′) increases with k. The mode $\tilde{q}$ of the extremal quotient is obtained from (3.4) For $k > 1$ this leads to

$$\tilde{q}^k = \frac{k - 1}{k + 1} < 1. \tag{3.5}$$

For $k \leqq 1$ no mode exists, and the distribution diminishes with q. The larger k, the smaller is the distance from the median to the mode, and hence, the smaller the asymmetry. The density of probability of the mode increases with k, and the probability

$$H_k(\tilde{q}) = \tfrac{1}{2}(1 - 1/k) \tag{3.6}$$

approaches $\frac{1}{2}$. The distribution (3.4) belongs to the Pareto type and has no moments of an order equal to or greater than k.

In N samples of sufficiently large size n, the largest quotient q_N^k, defined in the same way as u in equation (2.2), obtained from (3.4)

$$q_N^k = N - 1 \tag{3.7}$$

increases as a root of the number of samples, i.e. very quickly. The higher the order of the highest moments existing, the smaller will the expected largest quotient be.

From (3.4) and the symmetry (1.14) we obtain

$$H_k(q) - H_k(1/q) = 1 - 2/(1 + q^k). \tag{3.8}$$

The larger k, the larger is the percentage of the observations contained in the interval $1/q$ to q.

For a systematic estimate of k, the transformation (1.15) is used. Formula (3.4) leads to the probability $H^*(z)$ and the distribution $h^*(z)$ where

$$H^*(z) = \frac{1}{1 + e^{-kz}}; \qquad h^*(z) = \frac{ke^{-kz}}{(1 + e^{-kz})^2}. \tag{3.9}$$

The logarithm of the extremal quotient for initial distributions of the Cauchy type (where no moments of an order equaling or exceeding k exist) has the logistic distribution, [6], as the midrange $v = x_n + x_1$ for distributions of the exponential type (where all moments exist). The logarithm of the extremal quotient plotted on logistic probability paper should be scattered around a straight line

The order k of the lowest moment which diverges is obtained from the variance σ_z^2 of the distribution $h^*(z)$ which is [6]

$$\sigma_z^2 = \frac{\pi^2}{3k^2}. \tag{3.10}$$

For the estimate of k from (3.10), σ_z^2 is replaced by the estimate s_z^2 obtained from

$$s_z^2 = \frac{1}{N - 1} \sum_{\nu=1}^{N} \lg^2 \frac{x_{n,\nu}}{-x_{1,\nu}}. \tag{3.11}$$

For the Cauchy distribution itself, $k = 1$, and the probability and the distribution of the extremal quotient

$$H_1(q) = q/(1 + q); \qquad h_1(q) = (1 + q)^{-2}$$

are similar to the initial distribution.

The asymptotic distribution of the extremal quotient for initial distributions of the Cauchy type contains one parameter only, the order of the lowest diverging moment in the initial distribution All other traces of the initial distribution have disappeared.

4. Comparison of the extremal properties for the two types of initial distributions. Assume that the initial distribution is symmetrical, unlimited, and possesses an asymptotic distribution of the extremes. This is not always fulfilled. All moments may exist, and yet the distribution may not belong to the exponential type. No moments may exist, and yet the distribution may not belong to the Cauchy type. If the assumption holds, the initial distribution belongs either to the Cauchy, or to the exponential type.

We take N samples of size n, and estimate the median $\tilde{X}$ of the population from the central value m of the N central values of the samples Let $X_{1,v}$ and $X_{n,v}$ $(v = 1, 2, \cdots N)$ be the two extremes. If it happens for any v that

$$X_{1,v} > m \text{ or } X_{n,v} < m$$

the sample is too small, and its size has to be increased. The central value q of the observed extremal quotients $q_\nu = (X_{n,\nu} - m)/(m - X_{1,\nu})$ must be near unity.

If the initial distribution is of the exponential type, all moments in the population exist, and the midrange has the logistic distribution. If the initial distribution is of the Cauchy type, no moments of an order greater than k exist, and the logarithm of the extremal quotient has the logistic distribution. The order k can be estimated from the variance (3.11). If all moments in the population diverge, the calculation of the observed moments is futile since they do not characterize the population.

Addendum. The referee of this paper has suggested the following method for obtaining an asymptotic distribution of the extremal quotient for the exponential type. For large values of λ, formula (2.7) becomes, approximately,

$$H_\lambda(q) = \int_0^1 e^{-\lambda(z+zq)}\, d\lambda z.$$

Let

$$\lambda z = y.$$

Then

$$H_\lambda(q) = \int_0^\lambda \exp\left\{- y\left[1 + \left(\frac{y}{\lambda}\right)^{q-1}\right]\right\} dy.$$

The further transformation

$$e^t = \lambda^{q-1}, q-1 = t/lg\lambda,$$

leads to the probability $H^*(t)$ of the variate t

$$H^*(t) = \int_0^\lambda \exp\{-y[1 + e^{-t}y^{t/\lg\lambda}]\}\, dy,$$

whence asymptotically for $\lambda \to \infty$

$$H^*(t) = \int_0^\infty \exp\{-y(1 + e^{-t})\}\, dy$$

$$= 1/(1 + e^{-t}).$$

Therefore the logistic distribution holds at the same time for both initial types, using the transformation $t = \alpha u(q - 1)$ for the exponential type, and the logarithmic transformation for the Cauchy type.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[2] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the smallest and the largest member of a sample," *Proc. Camb. Philos. Soc.*, Vol. 24 (1928), p. 180.

[3] M. Fréchet, Sur la loi de probabilité de l'écart maximum. *Annales Soc. Polon. Math.*, Vol. 6 (1927).

[4] R. D. Gordon, "Values of Mills ratio of area to boundary ordinate and of the normal probability integral for large values of the argument," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 364–366.

[5] E. J. Gumbel, "The return period of flood flows," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 163–190.

[6] E. J. Gumbel, "Ranges and midranges," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 414–422.

[7] E. J. Gumbel, "On the independence of the extremes in a sample," *Annals of Math. Stat.*, Vol 17 (1946), pp. 78–81.

[8] R. von Mises, "La distribution de la plus grande de n valeurs," *Revue Math. de l'Union Interbalkanique*, Vol. 1 (1936).

ON A PRELIMINARY TEST FOR POOLING MEAN SQUARES IN THE ANALYSIS OF VARIANCE[1]

By A. E. Paull

Grain Research Laboratory, Winnipeg

Summary. The paper describes the consequences of performing a preliminary F-test in the analysis of variance. The use of the 5% or 25% significance level for the preliminary test results in disturbances that are frequently large enough to lead to incorrect inferences in the final test. A more stable procedure is recommended for performing the preliminary test in which the two mean squares are pooled only if their ratio is less than twice the 50% point.

I. Introduction

The problem discussed in this paper is one of a large class involving preliminary tests of significance. Studies of this type have recently been made by Bancroft [1] and Mosteller [2]. Bancroft dealt with a preliminary test for homogeneity of two variances, and a test of a regression coefficient. Mosteller dealt with the problem of pooling means from two normal populations having the same known variance. The present problem is an extension of Bancroft's work from investigations of the bias and variance of an estimate of variance, to investigations of the consequences of using that estimate in performing a further test of significance.

The problem arises frequently in the analysis of variance. As a simple example, consider an experiment carried out to test the hypothesis that different laboratories in a district all determine the protein content of wheat without systematic differences between laboratories. Three laboratories are selected at random and each is requested to analyze ten samples of the same wheat, five on each of two days. The analysis of variance would be set up in one of two ways:

MODEL I

Source of variation	*d f*	*M S*
Between laboratories	2	v_3
Between days within labs	3	v_2
Within days	24	v_1

MODEL II

Source of variation	*d f*	*M S*
Between laboratories	2	v_3
Within laboratories	27	$\dfrac{3v_2 + 24v_1}{27}$

The soundest procedure is to follow Model I in which the F-ratio, v_3/v_2, provides a valid though not very powerful test of the null hypothesis. But the investigator often doubts that this is the most effective form of analysis His past experience may have shown that measurements of this kind seldom exhibit day-to-day variations appreciably greater than their within-day variations. If he is willing to accept this credible assumption, he adopts Model II because

[1] Based on a doctoral dissertation submitted to the Faculty of North Carolina State College of the University of North Carolina at Raleigh, N. C, in June, 1948. Published as Paper No 107 of the Grain Research Laboratory, Board of Grain Commissioners, Winnipeg.

this increases the degrees of freedom from 2 and 3 to 2 and 27. These two models may conveniently be called the "never pool" and the "always pool" procedures.

The investigator often prefers what may be called a "sometimes pool" procedure. He starts with Model I and examines the null hypothesis that the variation between days is no greater than the variation within days by testing the F-ratio v_2/v_1. For this test, he selects a probability level P_1 that may be the 5% or some higher level. If the hypothesis of this preliminary test is not rejected, his judgement has been substantiated and he adopts Model II and pools the two mean squares. If the hypothesis is rejected, he retains Model I since he concludes that v_2 alone is the only valid estimate of error.

The following notation is introduced:

Degrees of freedom	*Mean square*	*Expected value of mean square*
n_3	v_3	σ_3^2
n_2	v_2	σ_2^2
n_1	v_1	σ_1^2

where $\sigma_1^2 \leq \sigma_2^2 \leq \sigma_3^2$.

The mean squares v_1, v_2, and v_3 are assumed to be distributed as central chi-squares. This assumption is justified if the treatments (laboratories in the example) are selected at random from a population of treatments. But if, as is more frequently the case, the experimenter is interested only in specified treatments, the non-central chi-square model is the appropriate one. However, if the two cases are sufficiently parallel, as seems probable, conclusions drawn from the central model may be expected to apply to the non-central model.

Let $\theta_{21} = \sigma_2^2/\sigma_1^2$ and $\theta_{32} = \sigma_3^2/\sigma_2^2$, and let $F(\nu_1, \nu_2, P)$ denote the value exceeded by F for ν_1 and ν_2 degrees of freedom with probability P. The rule of procedure for the "sometimes pool" test may be restated as follows:

Reject the main hypothesis that $\sigma_3^2 = \sigma_2^2(\theta_{32} = 1)$ if

$$v_2/v_1 \geq F_1(n_2, n_1; P_1) \quad \text{and} \quad v_3/v_2 \geq F_2(n_3, n_2; P_2)$$

or if

$$v_2/v_1 < F_1(n_2, n_1; P_1) \quad \text{and} \quad (n_2 + n_1)v_3/(n_2v_2 + n_1v_1) \geq F_3(n_3, n_2 + n_1; P_3).$$

The "never pool" procedure in which P_2 is used, and the "always pool" procedure in which P_3 is used, may be considered as special cases of the "sometimes pool" procedure in which P_1 takes on its extreme values, 1 and 0 respectively. In practice, the probability levels P_2 and P_3 are usually the same; in the present study they are allowed to be different in case this greater flexibility should prove desirable. The objects of the investigation are: (a) to examine the Type I error under the above rule of procedure, i.e., to determine the frequency of rejecting the null hypothesis when it is true; and (b) to examine the behaviour of the power with particular reference to comparisons with the power of the "never pool" procedure.

The remainder of this paper is divided into four sections: Part II contains a

general discussion of the results, conclusions and recommendations; and Part III illustrates the general conclusions with numerical examples The derivation of distributions, proofs by elementary arguments of general qualitative results, and derivations of closed form expressions for $n_3 = 2$, are given in Part IV.

II General Discussion of Results, Conclusions and Recommendations

2.1. Criterion employed. In this part the principal results and recommendations are discussed for the reader who is not interested in the mathematical details. To give results in a simple form is not easy, because of the many variables —the P's, the θ's, and the n's—that enter into the problem. It may be helpful to consider what is wrong with the "always pool" test, and then to state the properties which the preliminary test must have if it is to be regarded as useful and successful.

If the "always pool" procedure is employed when in fact σ_2^2 is greater than σ_1^2, i.e. $\theta_{21} > 1$, the denominator in the final F test tends to be too small. Thus the final F test gives too many significant results when its null hypothesis is true and if θ_{21} is great enough, there is no bound to this hidden distortion of the significance level. A test which the research worker thinks is being made at the 5% level might actually be at, say, the 47% level.

The preliminary test represents an attempt to avoid this alarming disturbance, since if θ_{21} is very large the test is expected to warn against pooling Such a procedure, however, can not be expected to remove this disturbance completely, and it does not do so, but to be successful it should keep the true or effective significance level of the final F test close to the nominal level at which the research worker thinks he is working.

A second requirement is that the preliminary test should increase the power in the final F test relative to the power of the "never pool" test When the powers of the "sometimes pool" and "never pool" tests are compared, it is important to make the comparison *at the same significance level.* Suppose the preliminary test shifts the significance level of the final F test from the 5% to the 6% level—a disturbance that for some uses would not be regarded as serious In this event the "sometimes pool" test (at the 6% level) would tend to be more powerful than the "never pool" test at the 5% level, because an increase in significance level generally results in an increase in power. But unless the "sometimes pool" test has more power than a "never pool" test made also at the 6% level, it has no real advantage over the "never pool" procedure

2.2. Effect of preliminary tests made at the 5% level. Probably the most common procedure in practice is to perform the preliminary test at the 5% level (i.e. $P_1 = .05$) and, whether pooling is prescribed or not, to conduct the final F test also at the 5% level, (i.e. $P_2 = P_3 = .05$). Such a procedure, except when θ_{21} is near one and the null hypothesis is true, results in the null hypothesis being rejected more frequently than if pooling is never resorted to

When the ratio θ_{21} is equal to one, so that routine pooling would be valid, the

preliminary test is effective. The true significance level of the final F test is *decreased* slightly, but is always confined between the 5% and the 4.75% levels. Further, the power is always greater than that of the "never pool" test made at the same significance level.

As θ_{21} increases from 1, the true significance level of the final F test increases to a maximum and then slowly decreases to 5%. Unfortunately the maximum need not be near to 5%: in the example presented later it is about 15%, and for a broad range of values of θ_{21} the true significance level is higher than 10%. Comparison with the power of the "never pool" test is also unfavorable to the "sometimes pool" test. For values of θ_{21} near 1, the "sometimes pool" test has the higher power, but as θ_{21} becomes larger the advantage passes to the "never pool" test.

When θ_{21} is very large there is, as would be expected, little disturbance. The preliminary test seldom prescribes pooling, so that the properties of the "sometimes pool" test are very similar to those of the "never pool" test, although the "never pool" procedure yields the slightly higher power.

The main objection to the use of the "sometimes pool" test is associated with the intermediate values of θ_{21}. If over a series of experiments θ_{21} has a moderate value greater than one, the "sometimes pool" test at the 5% levels yields more apparently significant results than are anticipated, and is also less powerful than a corresponding "never pool" test. The magnitude of these undesirable properties can be reduced somewhat by increasing the significance level of the preliminary test.

2.3. Effect of preliminary tests made at the 25% level. Use of the 25% instead of the 5% significance level for the preliminary test reduces, in general, the probability of rejecting the hypothesis. This reduction, at intermediate values of θ_{21}, results in a reduction of the extreme disturbances. When the ratio θ_{21} is equal to one, however, the effects are not as favourable. If the hypothesis is true, still fewer apparently significant results occur. A final test being carried out at the 5% level can now have an effective significance level close to 3.75%. If the hypothesis is false, the test is still more powerful than a corresponding "never pool" test but the gain is not as great as when a preliminary test at the 5% level is employed. Since most experimenters desire a reasonable amount of protection against an error in judgement of the true value of θ_{21}, the reduction in disturbances for intermediate values of θ_{21}, resulting from the use of the 25% rather than the 5% level, would be judged to outweigh the disadvantages of the compensating factors.

2.4. Effect of further increases in significance level. Increasing P_1, the significance level of the preliminary test, decreases the probability of rejecting the hypothesis only to the point where a critical value $\bar{P}_1$ is reached. Increasing P_1 beyond this value results in an increase in the probability of rejection. The properties of a "sometimes pool" test in which P_1 is less than $\bar{P}_1$ differ, in general, from those of a test in which P_1 is greater than $\bar{P}_1$.

Tests of the former type, which are referred to here as Class A tests, are the tests commonly encountered in practice Considering, for example, a test in which $P_2 = P_3 = .05$ and $n_1 = 20$, $n_2 = 4$, $n_3 = 2$, we find the critical value $\bar{P}_1$ to be .77, a figure much larger than the values .05 or .25 customarily chosen for P_1. The major portion of the present discussion deals with Class A tests. Tests in which P_1 is greater than $\bar{P}_1$ are referred to as Class B tests and discussion of their properties is relegated to a later section. An expression for evaluating $\bar{P}_1$ is given in Subsection 4. 3.

2.5. Effect of P_2, P_3. The probability levels (P_2, P_3) used for the final test determine the properties of the "sometimes pool" test for extreme values of θ_{21}. When θ_{21} is equal to one, the effective significance level is less than the nominal value P_3, but is not less than $(1 - P_1)P_3$. The power of such a test is greater than the power of a corresponding "never pool" test, but less than the power of a test in which one always pools and uses the P_3 level. For very large values of θ_{21} the behavior of the "sometimes pool" test approaches, in all respects, the behaviour of a "never pool" test at the P_2 level.

2.6. Effect of n_2, n_1. The degrees of freedom n_2 and n_1, associated with the mean squares that are sometimes pooled, clearly affect the magnitude of the disturbance. Because analytic investigation becomes complex, the following remarks are based on conjectures arising out of examination of a number of numerical examples.

A large value of n_2 is desirable in two respects. As n_2 becomes larger the preliminary test becomes more powerful and pooling is prescribed less often. In addition, when pooling is prescribed the pooled mean square is further weighted in favour of the valid error σ_2^2. Both factors are contributing towards a decrease in bias of the error mean square with a consequent reduction in the disturbance introduced into the final test.

The effect of n_1 is not as simple. As n_1 becomes larger the preliminary test again becomes more powerful and pooling is prescribed less often. But when pooling is prescribed, the pooled mean square in this case is further weighted in favour of σ_1^2, which is smaller than the valid error σ_2^2. The effect on the final test, which is due to a combination of these two factors, clearly depends on the value of θ_{21}. For intermediate values of θ_{21} the latter factor is the predominant one, and the disturbance of the effective significance level is increased as n_1 is increased.

2.7. Class B Test. A Class B test is one in which the probability level (P_1) of the preliminary test is greater than a critical value $\bar{P}_1$. Pooling is prescribed only when the mean square v_1 is relatively large, with the result that the error mean square tends to be too large. Accordingly, a Class B "sometimes pool" test rejects the hypothesis less frequently than a "never pool" test at the P_2 level

The effective significance level of a Class B test is less than P_2 for all values of θ_{21}. It has its lowest value when θ_{21} is equal to one, and approaches P_2 as θ_{21}

becomes very large. Because pooling is prescribed infrequently, little power is gained by the use of a Class B test rather than a "never pool" test.

2.8. Recommendations. The principal conclusions discussed in the preceding subsections may be summarized as follows: A preliminary test carried out at a significance level as low as 5% affords little protection against errors in judgement. If σ_1^2 is equal to $\sigma_2^2(\theta_{21} = 1)$ the reduction in errors of inference is appreciable; but if, in fact, σ_1^2 is less than $\sigma_2^2(\theta_{21} > 1)$, a greater number of incorrect inferences are made than if a preliminary test is not employed at all. The use of the 25% significance level for the preliminary test introduces the same disturbances but to a lesser extent. Extreme increases in the effective significance level at possible values of θ_{21} are reduced and losses in power at these values are not as serious. The 25% level provides a reasonable amount of protection against an error in judgement regarding the true value of θ_{21}. However, when n_2 is large relative to n_1, a smaller significance level could be employed without introducing any serious disturbances at the intermediate values of θ_{21}, and with a resulting gain in power at values of θ_{21} near one.

The following method of performing a preliminary test is recommended as one which tends to stabilize the disturbances at intermediate values of θ_{21} while still taking advantage of a considerable portion of the possible gain in power at values of θ_{21} near one. The procedure consists of pooling the two mean squares v_2 and v_1 only if their ratio is less than 2 F_{50}, where F_{50} is the 50 per cent point of the F-distribution for n_2 and n_1 degrees of freedom. The use of the multiple 2 is arbitrary and a smaller value may be used if the experimenter desires additional control over extreme disturbances.

This procedure has the advantage of admitting less disturbance over a larger range of values of n_2 and n_1. The customary method prescribes pooling if the null hypothesis ($\theta_{21} = 1$) of the preliminary test is not rejected at some preassigned probability level P_1. If enough observations are available to provide reliable values for v_2 and v_1, pooling is prescribed only if σ_2^2 and σ_1^2 are essentially the same. However, if small numbers of degrees of freedom are involved, the preliminary test is too weak to reject the hypothesis even if σ_1^2 is appreciably less than σ_2^2, and pooling will be prescribed too frequently. On the other hand, the use of the recommended procedure has the effect of prescribing pooling only when it can be said, with confidence exceeding 50%, that the true value of θ_{21} is less than some chosen value such as 2.

This can be demonstrated simply by considering a series of experiments in which preliminary tests are performed. When $v_2/v_1 < 2F_{50}$, we make the statement

$$\theta_{21} < 2, \tag{1}$$

and when $v_2/v_1 \geq 2F_{50}$, we make the statement

$$\theta_{21} \geq 2. \tag{2}$$

We have

$$Pr\left\{\frac{v_2}{v_1} \cdot \frac{1}{\theta_{21}} \geq F_{50}\right\} = .50,$$

or

$$Pr\left\{\frac{v_2}{v_1} \geq F_{50}\,\theta_{21}\right\} = .50.$$

If statement (1) is true,

$$Pr\left\{\frac{v_2}{v_1} < 2F_{50}\right\} \geq .50;$$

and if statement (2) is true,

$$Pr\left\{\frac{v_2}{v_1} \geq 2F_{50}\right\} \geq 50.$$

Thus, no matter what the true value of θ_{21}, the statements are true more than 50% of the time.

Fifty per cent points of the F-distribution have been tabulated by Merrington and Thompson [3].

A simpler rule, and one which is nearly equivalent when the degrees of freedom involved are each greater than 6, is to pool if the ratio of the mean squares is less than 2, without any reference to the F-table. For smaller numbers of degrees of freedom, however, this simpler rule does not embody the advantages of the $2F_{50}$ rule, unless of course, n_1 and n_2 are equal.

III. Numerical Illustrations

3.1. Effect of P_1 illustrated. An example of the influence of P_1 on the effective significance level or Type I error of a "sometimes pool" test is illustrated in Figure 1. When $P_1 = 0$, the Type I error has its maximum value equivalent to the Type I error of an "always pool" test at the P_3 level. As P_1 increases from zero, the Type I error decreases until at $P_1 = \bar{P}_1$(.77 in this case) it reaches its minimum value at a level less than P_2. As P_1 increases from $\bar{P}_1$, the Type I error increases until, at $P_1 = 1$, the Type I error is equal to P_2

The influence of P_1 on the power of a "sometimes pool" test is illustrated in Figure 2. The gain in power, as a function of θ_{21}, is presented for three Class A tests. Since comparisons of power are made over tests having different Type I errors, the gain is expressed as the proportion actually attained of the total gain in power that is possible if the true value of θ_{21} is actually known. When $P_1 = \bar{P}_1 = .77$, the curve is observed to decrease monotonically to zero. However, for lower values of P_1, the preliminary test prescribes pooling more often, and more power is gained when θ_{21} is near one but less power is gained or power is actually lost when θ_{21} is large

The power gained or lost at various values of θ_{21} is illustrated in Table I. The probability of rejecting the hypothesis for the "sometimes pool" test is

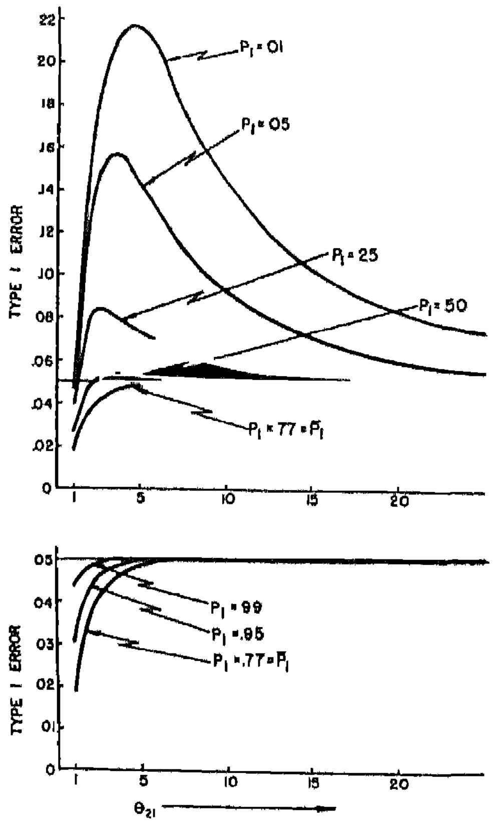


FIG. 1. Effect of Varying P_1. $n_1 = 20$, $n_2 = 4$, $n_3 = 2$ and $P_2 = P_3 = .05$. (a) Upper diagram: Class A Tests (b) Lower diagram. Class B Tests.

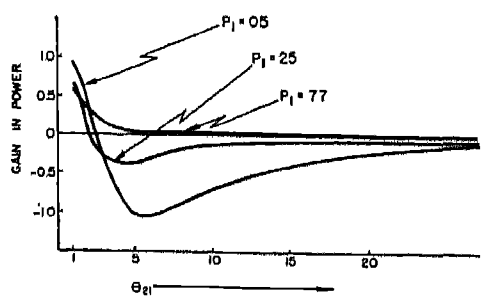


FIG. 2. Proportion of Possible Gain in Power Actually Attained. $n_1 = 20$, $n_2 = 4$, $n_3 = 2$, $P_2 = P_3 = .05$.

tabulated opposite "s p.", and for the "never pool" test *having the same Type I error* opposite "n p.".

The last line of the table approaches the probabilities for a "never pool" test having a Type I error of 5%. Except for values very near $(\theta_{21}, \theta_{32}) = (1, 1)$, the probability of rejecting the null hypothesis, using a "sometimes pool" test, is greater than if a "never pool" test, *at the* 5% level is used. In this sense, the

TABLE I

Comparison of Power of a "Sometimes Pool" (s p.) Test and Corresponding "Never Pool" (n.p.) Tests

$n_1 = 20,\ n_2 = 4,\ n_3 = 2;\ P_1 = P_2 = P_3 = .05$

Value of θ_{21}	Test	Type I Error $\theta_{32} = 1$	Value of θ_{32}							
			1 8	2 8	4.3	7.1	12.5	25	50	750
1.0	s.p.	.048	164	.299	443	599	739	.855	922	.984
	n.p.	.048	.112	.192	.297	.441	604	765	.870	.972
1 2	s p.	.067	200	.338	476	621	751	.860	925	.984
	n.p.	067	.149	245	.361	.508	662	805	895	978
1 6	s p.	102	248	379	.503	.632	.750	.855	.921	.983
	n.p	102	.210	323	447	.592	730	849	920	.983
2 0	s.p.	.127	.271	390	.500	.619	736	845	915	.981
	n.p.	.127	.250	.370	.497	.636	.764	.870	932	.986
2.5	s p.	146	.278	.382	482	596	715	.831	907	.975
	n.p	.146	.278	.402	528	.664	784	.882	938	.987
4 5	s.p.	148	.233	.309	399	.520	.657	796	.887	.976
	n.p.	.148	280	.405	531	.666	786	883	939	.987
7.0	s.p.	.117	.182	.255	350	.482	632	.781	880	.974
	n.p.	.117	.234	352	.478	.620	.751	862	.927	985
10	s.p.	.091	.152	.227	327	465	621	.776	.877	.974
	n p.	.091	.191	.300	422	569	712	.838	.913	.982
16	s.p	.067	.130	209	313	456	615	773	.875	.973
	n.p.	.067	.149	.245	361	509	662	.805	.895	.978
100	s.p.	.051	117	.200	.307	.452	.613	.771	.875	.973
	n.p.	.051	.118	.201	.308	454	615	773	875	973

Below the heavy line the s p test is less powerful then the n p. test.

"power" of the "sometimes pool" test is greater everywhere except near $(\theta_{21}, \theta_{32}) = (1, 1)$.

3.2. Effect of P_2, P_3 illustrated. The influence of the probability levels employed in the final phase of a "sometimes pool" test is illustrated in Figure 3. The main effect is observed to be the manner in which the behaviour is constrained at the extreme values of θ_{21}.

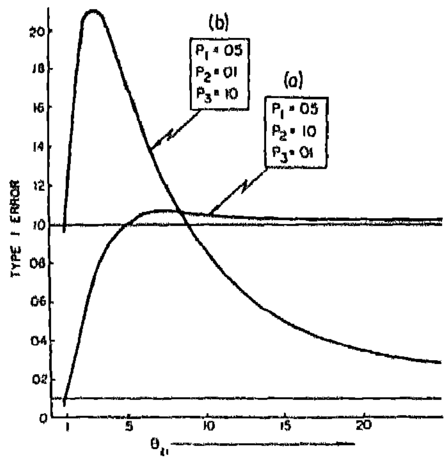


FIG. 3. Class A Tests, $n_1 = 20$, $n_2 = 4$, $n_3 = 2$.

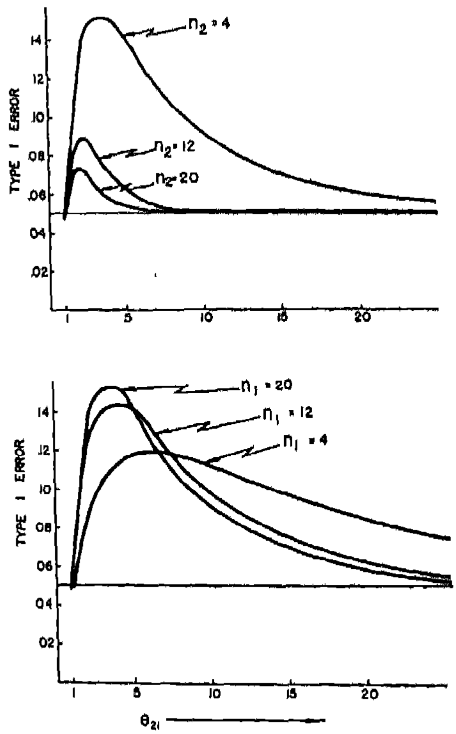


FIG. 4. (a) Upper Diagram: Effect of Varying n_2. $P_1 = P_2 = P_3 = .05$ and $n_1 = 20$, $n_3 = 2$. (b) Lower Diagram: Effect of Varying n_1 $P_1 = P_2 = P_3 = .05$ and $n_2 = 4$, $n_3 = 2$.

3.3. Effect of n_2, n_1 illustrated. The response of the Type I error to increases in the degrees of freedom of the preliminary test is illustrated in Figure 4. The maximum disturbance is observed to increase as n_1 increases or as n_2 decreases.

3.4. Class B test illustrated. The behaviour of the Type I error of some Class B tests is illustrated in Figure 1(b). The hypothesis is always rejected less frequently than if a "never pool" test at the P_2 level is used.

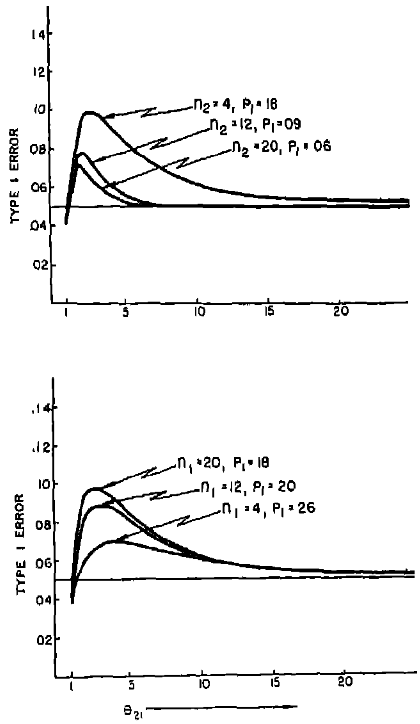


FIG. 5. (a) Upper Diagram: Effect of Varying n_2 when $F_1 = 2F_{30}$, $P_2 = P_3 = .05$ and $n_1 = 20$, $n_3 = 2$. (b) Lower Diagram. Effect of Varying n_1 when $F_1 = 2F_{50}$, $P_2 = P_3 = .05$ and $n_2 = 4$, $n_3 = 2$.

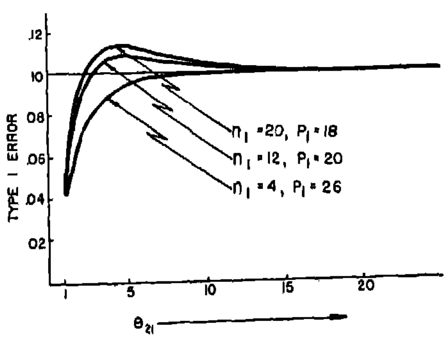


FIG. 6. Effect of Varying n_1 when $P_2 > P_3$. $P_2 = .10$, $P_3 = .05$ and $n_2 = 4$, $n_3 = 2$.

3.5. Recommended procedure illustrated. Figure 5 illustrates the behaviour of the Type I error when the recommended procedure is applied to the special cases presented in Figure 4. When $n_1 = 12$, $n_2 = 4$, the 20% probability level is

prescribed and the Type I error never exceeds .09. When $n_1 = 20$, $n_2 = 20$, the more liberal value of 6% is prescribed and the resulting Type I error never exceeds .07. The more liberal choice of P_1 results in a greater gain of power, near $\theta_{21} = 1$, than would have resulted if the 20% level had been used throughout. A small loss in power occurs when θ_{21} is large. Should the experimenter wish to guard against this loss in power for a larger range of values of θ_{21} near one, he may do so, at the expense of a somewhat larger disturbance in the Type I error, by choosing P_2 larger than P_3. In the present example, if P_2 is taken as .10 instead of .05, Figure 6 shows that the Type I error is changed only slightly for values of θ_{21} near one, but the maximum disturbance is increased. Such a test, is uniformly more powerful than the "never pool" test for all values of θ_{21} for which the Type I error is less than .10; a much larger range of values than in the previous case.

IV. Derivations and Proofs

4.1. Derivation of joint frequency function. The joint frequency function of the v's is given by

$$c_1 v_1^{\frac{1}{2}n_1-1} v_2^{\frac{1}{2}n_2-1} v_3^{\frac{1}{2}n_3-1} \exp\left\{-\tfrac{1}{2}\left[\frac{n_1 v_1}{\sigma_1^2} + \frac{n_2 v_2}{\sigma_2^2} + \frac{n_3 v_3}{\sigma_3^2}\right]\right\},$$

where c_1 is independent of the v's. Transform to new variables:

$$u_1 = \frac{n_2 v_2}{n_1 v_1}, \qquad u_2 = \frac{n_3 v_3}{n_2 v_2}, \qquad w = \frac{n_1 v_1}{n_3}.$$

By integrating and evaluating the constant, the joint frequency function of u_1 and u_2 is obtained:

$$(3) \qquad p = \frac{\theta_{21}^{\frac{1}{2}n_1}\,\theta_{32}^{\frac{1}{2}(n_2+n_1)}}{\mathrm{B}(\frac{1}{2}n_2, \frac{1}{2}n_1)\mathrm{B}(\frac{1}{2}n_3, \frac{1}{2}(n_1+n_2))} \cdot \frac{u_1^{\frac{1}{2}(n_3+n_2)-1}\, u_2^{\frac{1}{2}n_3-1}}{(\theta_{21}\theta_{32} + \theta_{32}u_1 + u_1u_2)^{\frac{1}{2}(n_3+n_2+n_1)}}$$

where $\theta_{21} = \sigma_2^2/\sigma_1^2$; $\theta_{32} = \sigma_3^2/\sigma_2^2$.

4.2. Definition of critical region. The rule of procedure for the "sometimes pool" test may now be expressed in terms of the u's. Reject the hypothesis $\theta_{32} = 1$ if

$$\begin{cases} u_1 \geq u_1^0, \\ u_2 \geq u_2^0, \end{cases} \quad \text{or} \quad \begin{cases} u_1 < u_1^0, \\ \dfrac{u_1 u_2}{1 + u_1} \geq u_3^0, \end{cases}$$

where

$$u_1^0 = \frac{n_2}{n_1} \cdot F_1(n_2, n_1; P_1),$$

$$u_2^0 = \frac{n_3}{n_2} \cdot F_2(n_3, n_2; P_2),$$

$$u_3^0 = \frac{n_3}{n_2 + n_1} \cdot F_3(n_3, n_2 + n_1; P_3).$$

The reader will note that the u's are ratios of sums of squares. The symbol u_1 is associated with the preliminary test The final test when pooling is not prescribed is associated with the symbol u_2, and when pooling is prescribed the relevant statistic is $u_1u_2/(1 + u_1)$.

The critical region defined in this way is illustrated in the two dimensional sample space $\{u_1, u_2\}$ of Figure 7(a). The critical regions of the "never pool" and the "always pool" test are readily identified in this figure. The region of a "never pool" test at the P_2 level is designated by $A + B_1 + C$, the area above the line $u_2 = u_2^0$; and the region of an "always pool" test at the P_3 level is designated by $B_1 + B_2 + C + D$, the area above the curve $u_1u_2 = u_3^0(1 + u_1)$. The critical region of the "sometimes pool" test, $B_1 + B_2 + C$, may be considered in two parts: the portion due to *pooling*, $B_1 + B_2$, and the portion due to *not pooling*, C.

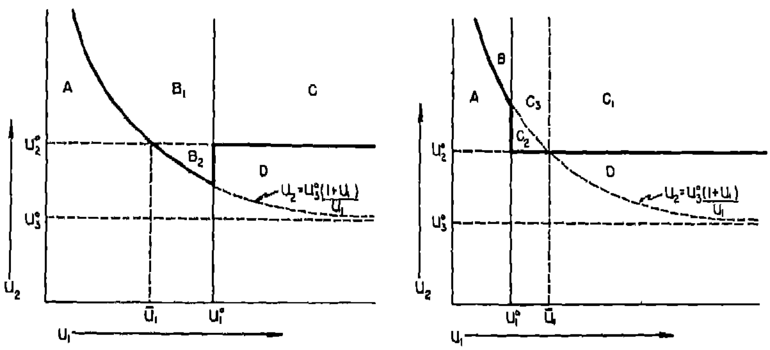


FIG. 7. Critical Region of "Sometimes Pool" Test. (a) Left: Class A Test: $u_1^0 > \bar{u}_1$ (b) Right: Class B Test: $u_1^0 < \bar{u}_1$.

The probability of rejecting the null hypothesis is given by

$$Q(\theta_{21}, \theta_{32}) = \int_0^{u_1^0}\int_w^\infty p\, du_1\, du_2 + \int_{u_1^0}^\infty\int_{u_2^0}^\infty p\, du_1\, du_2, \tag{4}$$

where p is the frequency function (3), and $w = u_3^0(1 + u_1)/u_1$.

Simple explicit expressions for these integrals cannot be obtained in general, but when $n_3 = 2$ they can be reduced to forms containing incomplete beta functions. This special case is dealt with in Subsection 4.7.

4.3. Critical value of P_1. The symbol $\bar{u}_1$ in Figure 1 is used to denote the u_1 coordinate of the point of intersection of the line $u_2 = u_2^0$ and the curve $u_1u_2 = u_3^0(1 + u_1)$. Accordingly,

$$\bar{u}_1 = \frac{u_3^0}{u_2^0 - u_3^0}, \tag{5}$$

a value readily determined for any given test This relationship may be expressed in terms of the F's as

$$\bar{F}_1 = \frac{1}{\frac{n_2}{n_1}\left\{\frac{F_2}{F_3} - 1\right\} + \frac{F_2}{F_3}}, \tag{6}$$

NATI

where $\bar{F}_1$ is defined by $n_1\bar{u}_1 = n_2\bar{F}_1$. The probability level corresponding to $\bar{F}_1$ is denoted by $\bar{P}_1$.

The critical value $\bar{P}_1$ is the value of P_1 which divides the possible "sometimes pool" tests into two types having different properties. If P_1 is less than $\bar{P}_1(F_1 > \bar{F}_1$ or $u_1^0 > \bar{u}_1)$, the test is referred to as a Class A test. If P_1 is greater than $\bar{P}_1(F_1 < \bar{F}_1$ or $u_1^0 < \bar{u}_1)$, the test is referred to as a Class B test.

4.4. Lemma 1.

LEMMA 1. *If $\theta_{21}' \geq \theta_{21}$ and $\theta_{32}' \geq \theta_{32}$, and if the equality applies in one of these, then the ratio of the frequency functions* (3)

$$\frac{p(u_1, u_2 \mid \theta_{21}', \theta_{32}')}{p(u_1, u_2 \mid \theta_{21}, \theta_{32})} \tag{7}$$

increases monotonically as (i) *u_1 increases with u_2 fixed, or as* (ii) *u_2 increases with u_1 fixed, or as* (iii) *u_1 increases on fixed pooling curve $u_1u_2 = u_3^0(1 + u_1)$.*

PROOF. The ratio (7) is a monotonic function of

$$\frac{\theta_{21}\theta_{32} + \theta_{32}u_1 + u_1u_2}{\theta_{21}'\theta_{32}' + \theta_{32}'u_1 + u_1u_2}.$$

It is easily shown that an expression of the form $(a + bx)/(c + dx)$ increases monotonically with respect to x if $a/c < b/d$, and this condition holds for cases (i), (ii), and (iii).

4.5. Lemma 2.

LEMMA 2. *If area L lies above a given pooling curve, and to the right of a given preliminary line, if area K lies below the same pooling curve, and to the left of the same preliminary line, and if*

$$Pr\{L \mid \theta_{21}, \theta_{32}\} \geq Pr\{K \mid \theta_{21}, \theta_{32}\},$$

then

$$Pr\{L \mid \theta_{21}', \theta_{32}'\} > Pr\{K \mid \theta_{21}', \theta_{32}'\},$$

where $\theta_{21}' \geq \theta_{21}$ and $\theta_{32}' \geq \theta_{32}$ and the equality applies in one of these.

PROOF. For any point (u_1, u_2) in K and any point (u_1', u_2') in L, Lemma 1 (iii) yields

$$\frac{p(u_1, u_2 \mid \theta_{21}', \theta_{32}')}{p(u_1, u_2 \mid \theta_{21}, \theta_{32})} < \frac{p(u_1', u_2'' \mid \theta_{21}', \theta_{32}')}{p(u_1', u_2'' \mid \theta_{21}, \theta_{32})},$$

where $u_2'' = c(1 + u_1')/u_1'$, and c is a constant defined by $u_2 = c(1 + u_1)/u_1$. Since K is below a given pooling curve, $u_2'' < u_2'$ and

$$\frac{p(u_1', u_2'' \mid \theta_{21}', \theta_{32}')}{p(u_1', u_2'' \mid \theta_{21}, \theta_{32})} < \frac{p(u_1', u_2' \mid \theta_{21}', \theta_{32}')}{p(u_1', u_2' \mid \theta_{21}, \theta_{32})}.$$

Consider

$$\frac{p(u_1, u_2 \mid \theta'_{21}, \theta'_{32})}{p(u_1, u_2 \mid \theta_{21}, \theta_{32})} < b < \frac{p(u'_1, u'_2 \mid \theta'_{21}, \theta'_{32})}{p(u'_1, u'_2 \mid \theta_{21}, \theta_{32})},$$

where b is a constant such that the inequalities hold for all (u_1, u_2) in K and all (u'_1, u'_2) in L.

Integrating over the regions yields

$$Pr\{K \mid \theta'_{21}, \theta'_{32}\} < b.Pr\{K \mid \theta_{21}, \theta_{32}\}$$

and

$$b.Pr\{L \mid \theta_{21}, \theta_{32}\} < Pr\{L \mid \theta'_{21}, \theta'_{32}\}.$$

But

$$Pr\{K \mid \theta_{21}, \theta_{32}\} \leq Pr\{L \mid \theta_{21}, \theta_{32}\},$$

thus

$$Pr\{K \mid \theta'_{21}, \theta'_{32}\} < Pr\{L \mid \theta'_{21}, \theta'_{32}\},$$

which completes the proof.

4.6. General Properties.

RESULT 1. *When $\theta_{21} = 1$, the Type I error of a Class A test is less than P_3.*

PROOF. In the notation of Fig. 7(a), the probability of falling in $B_1 + B_2 + C + D$ is P_3 when $\theta_{21} = 1$ and $\theta_{32} = 1$. The region of rejection of the "sometimes pool" test is smaller by D.

RESULT 2. *When $\theta_{21} = 1$, the Type I error of a Class A test is greater than $(1 - P_1)P_3$.*

PROOF. The statistics u_1 and $u_1u_2/(1 + u_1)$ are independent when $\theta_{21} = 1$ and $\theta_{32} = 1$. Under these conditions, the probability of falling in $B_1 + B_2$, in the notation of Fig. 7(a), is equal to the product of two incomplete beta functions having the values $(1 - P_1)$ and P_3. Consequently, the Type I error is greater than $(1 - P_1)P_3$.

RESULT 3. *The Type I error approaches P_2 as θ_{21} approaches infinity.*

PROOF. The distribution becomes singular when $\theta_{21} = \infty$. The frequency function approaches zero uniformly for any finite value of u_1 and approaches

$$\frac{1}{B(\frac{1}{2}n_3, \frac{1}{2}n_2)} \frac{u_2^{\frac{1}{2}n_3 - 1}}{(1 + u_2)^{\frac{1}{2}(n_3 + n_2)}}$$

at $u_1 = \infty$. When $\theta_{21} = \infty$, the entire mass is concentrated on the line $u_1 = \infty$ and is distributed as a beta variable along that line. In the notation of Fig. 7(a), $Pr(B_1 + B_2) \to 0$ and $Pr(C) \to P_2$.

RESULT 4. *If the Type I error of a Class A test is Q_0 for θ_{21}, then for $\theta'_{21} > \theta_{21}$, the Type I error is greater than r, where r is equal to the lesser of Q_0 and P_2.*

Three useful corollaries are associated with the above result:

RESULT 4.1. *If at $\theta_{21} = 1$, the value of the Type I error is less than P_2, this is its minimum value for any θ_{21}.*

RESULT 4.2. *If at $\theta_{21} = 1$, the Type I error is less than P_2, then as θ_{21} increases from 1 the Type I error increases monotonically until P_2 is reached.*

RESULT 4.3. *If for some value of θ_{21} the Type I error is equal to or greater than P_2, then for any larger value of θ_{21}, the Type I error is greater than P_2.*

PROOF. Let the regions of Fig. 8 be denoted by $R_1 = A_1 + B_1 + C_1$ with similar designations for R_2 and R_3. Let $R_4 = B_1 + B_2 + B_3 + B_4 + C_1 + C_2$.

If $r = Q_0$, let the non-pooling line between R_1 and R_2 in Fig. 8 correspond to Q_0 for all θ_{21}. Then $Pr\{R_4 \mid \theta_{21}, 1\} = Pr\{R_1 \mid \theta_{21}, 1\}$, whence $Pr\{B_2 + B_3 + B_4 + C_2 \mid \theta_{21}, 1\} = Pr\{A_1 \mid \theta_{21}, 1\}$. By Lemma 2, we have for any $\theta_{21}' > \theta_{21}$, $Pr\{B_2 + B_3 + B_4 + C_2 \mid \theta_{21}', 1\} > Pr\{A_1 \mid \theta_{21}', 1\}$ and $Pr\{R_4 \mid \theta_{21}', 1\} > Pr\{R_1 \mid \theta_{21}', 1\} = Q_0$.

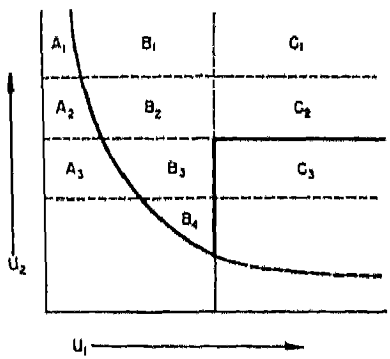


FIG. 8. Critical Regions for Result 4.

If $r = P_2$, let the non-pooling line at the lower boundary of R_3 in Fig. 8 correspond to Q_0 for all θ_{21}. Then in the same way $Pr\{B_4 \mid \theta_{21}, 1\} = Pr\{A_1 + A_2 + A_3 + C_3 \mid \theta_{21}, 1\}$ and $Pr\{B_4 \mid \theta_{21}', 1\} > Pr\{A_1 + A_2 + A_3 \mid \theta_{21}', 1\}$ by Lemma 2. Thus $Pr\{R_4 \mid \theta_{21}', 1\} > Pr\{R_1 + R_2 + A_3 + B_3 \mid \theta_{21}', 1\}$ and $Pr\{R_4 \mid \theta_{21}', 1\} > Pr\{R_1 + R_2 \mid \theta_{21}', 1\} = P_2$.

RESULT 5. *For a Class B test, the Type I error is less than P_2 for all θ_{21}.*

PROOF. Figure 7(b) illustrates the critical region of a Class B test. We have $Pr\{A + B + C_1 + C_2 + C_3\} = P_2$. But the region of rejection of the "sometimes pool" test is smaller, excluding A.

RESULT 6. *The Type I error of a Class B test, for $\theta_{21} = 1$, is greater than $(1 - \bar{P}_1)P_3$.*

PROOF. Changing P_1 to $\bar{P}_1$ removes C_2 from the region of rejection in Fig. 7(b), thus decreasing the Type I error. The modified test lies in both Class B and Class A, so that Result 2 applies.

RESULT 7. *For any θ_{21}, the Type I error is a minimum for changes of P_1 when $P_1 = \bar{P}_1$.*

PROOF. For a Class A test, changing P_1 to $\bar{P}_1$ removes region B_2 of Fig 7(a), thus decreasing the Type I error. For a Class B test, changing P_1 to $\bar{P}_1$ removes region C_2 of Fig. 7(b), similarly decreasing the Type I error.

RESULT 8. *A Class A test, in which the Type I error is less than or equal to P_2, is more powerful than a "never pool" test having the same Type I error.*

PROOF. In Fig. 8, let region $R_1 = A_1 + B_1 + C_1$ be equal in size to $R_4 = B_1 + B_2 + B_3 + B_4 + C_1 + C_2$. Then $Pr\{R_4 \mid \theta_{21}, 1\} = Pr\{R_1 \mid \theta_{21}, 1\}$ and $Pr\{B_2 + B_3 + B_4 + C_2 \mid \theta_{21}, 1\} = Pr\{A_1 \mid \theta_{21}, 1\}$ Increasing $\theta_{32} = 1$ to θ'_{32} and applying Lemma 2 yields $Pr\{R_4 \mid \theta_{21}, \theta'_{32}\} > Pr\{R_1 \mid \theta_{21}, \theta_{32}\}$.

RESULT 9. *For a fixed Type I error a Class A test, carried out at given levels of P_2 and P_3, is more powerful than a Class B test at the same levels.*

PROOF. Fig. 7 and Lemma 2 apply at once

4.7. Closed form expressions for $n_3 = 2$. The probability of rejecting the hypothesis in a "sometimes pool" test is given by $Q(\theta_{21}, \theta_{32}) = Q_1 + Q_2$ where Q_1 corresponds to the region B, and Q_2 to the region C of Fig. 7.

The integrals (4) representing the probability of rejecting the null hypothesis, reduce, when $n_3 = 2$, to

$$Q_1 = \left\{\frac{1 + \dfrac{u_3^0}{\theta_{32}}}{1 + \dfrac{u_3^0}{\theta_{21}\theta_{32}}}\right\}^{\frac{1}{2}n_1} \frac{I_z(\frac{1}{2}n_2, \frac{1}{2}n_1)}{\left\{1 + \dfrac{u_3^0}{\theta_{32}}\right\}^{\frac{1}{2}(n_2+n_1)}}, \tag{8}$$

where the argument z of the incomplete beta function is defined by $z = x/(1 + x)$ where

$$x = \left\{\frac{1 + \dfrac{u_3^0}{\theta_{32}}}{1 + \dfrac{u_3^0}{\theta_{21}\theta_{32}}}\right\}. \tag{9}$$

Under the null hypothesis $\theta_{32} = 1$,

$$Q_1 = I_z(\tfrac{1}{2}n_2, \tfrac{1}{2}n_1) \left\{\frac{1 + u_3^0}{1 + \dfrac{u_3^0}{\theta_{21}}}\right\}^{\frac{1}{2}n_1} \cdot P_3, \tag{10}$$

since

$$P_3 = \frac{1}{(1 + u_3^0)^{\frac{1}{2}(n_2+n_1)}}.$$

Similarly

$$Q_2 = \frac{I_{z'}(\frac{1}{2}n_1, \frac{1}{2}n_2)}{\left\{1 + \dfrac{u_2^0}{\theta_{32}}\right\}^{\frac{1}{2}n_2}}, \tag{11}$$

where the argument z' of the incomplete beta function is defined by $z' = 1/(1+x')$ where

$$x' = \left\{1 + \frac{u_2^0}{\theta_{.2}}\right\}\frac{u_1^0}{\theta_{21}}. \tag{12}$$

Under the null hypothesis $\theta_{.2} = 1$,

$$Q_2 = I_{z'}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2)\cdot P_2, \tag{13}$$

since

$$P_2 = \frac{1}{(1 + u_2^0)^{\frac{1}{2}n_2}}.$$

The incomplete beta function has been tabulated by Pearson [4].

The author wishes to thank Professor W. G. Cochran and Professor John W. Tukey for helpful advice in the preparation of this paper.

REFERENCES

[1] T. A. Bancroft. "On biases in estimation due to the use of preliminary tests of significance". *Annals of Math. Stat.*, Vol 15 (1944), pp. 190-204.

[2] Frederick Mosteller. "On pooling data". *Jour. Am. Stat Assn.*, Vol. 43 (1948), pp. 231-242.

[3] M. Merrington and C. M. Thompson. "Tables of percentage points of the inverted beta (*F*) distribution". *Biometrika*, Vol. 33 (1943), pp. 73-88.

[4] Karl Pearson, *Tables of the Incomplete Beta Function*, Cambridge University Press, 1934.

ESTIMATING THE MEAN AND VARIANCE OF NORMAL POPULATIONS FROM SINGLY TRUNCATED AND DOUBLY TRUNCATED SAMPLES[1]

BY A. C. COHEN, JR.

The University of Georgia

1. Summary. This paper is concerned with the problem of estimating the mean and variance of normal populations from singly and doubly truncated samples having known truncation points. Maximum likelihood estimating equations are derived which, with the aid of standard tables of areas and ordinates of the normal frequency function, can be readily solved by simple iterative processes. Asymptotic variances and covariances of these estimates are obtained from the information matrices. Numerical examples are given which illustrate the practical application of these results. In Sections 3 to 8 inclusive, the following cases of doubly truncated samples are considered: I, number of unmeasured observations unknown; II, number of unmeasured observations in each 'tail' known; and III[2], total number of unmeasured observations known, but not the number in each 'tail'. In Section 9, singly truncated samples are treated as special cases of I and II above.

2. Introduction. In practice, truncated samples arise with various types of experimental data in which recorded measurements are available over only a partial range of the variable Such samples are usually classified according to the form of the population (complete) distribution; according to whether the truncation points are known or unknown; and according to whether the number of unmeasured (missing) observations is known or unknown In this paper, the further classification of singly truncated or doubly truncated is made, accordingly as one or both 'tails' of the sample have been removed. Pearson and Lee [1, 2], Fisher [3], Hald [4][3], and this writer [5] studied singly truncated normal samples with a known truncation point when the number of unmeasured observations is unknown. Stevens [6], Cochran [7], and Hald [4] studied similar samples with a known number of unmeasured observations. Stevens [6] also considered doubly truncated normal samples with known truncation points when the number of unmeasured observations in each 'tail' is known. In each of these papers, equations were derived with which maximum likelihood estimates of the population mean and variance can be computed from samples of the type considered. With the exception of [5], which uses standard tables of the normal frequency

[1] Based on papers presented before the American Mathematical Society, Durham, North Carolina, April 2, 1949, and before a joint meeting of the Institute of Mathematical Statistics and the Biometric Society, Chapel Hill, North Carolina, March 18, 1950

[2] The problem involved in this case was recently called to the writer's attention by Churchill Eisenhart

[3] Reference [4] appeared while this paper was awaiting publication. Minor revisions have been made in view of Hald's results.

function, practical application of the various estimating equations involves use of special tables which may frequently be unavailable.

3. Case I. Number of unmeasured observations unknown. Let x_0' designate the left truncation point, $x_0' + R$ the right truncation point, and hence R the sample range. Let n_0 be the number of measured observations with values equal to or between the truncation points. In this case, the number of unmeasured observations is assumed to be unknown. We translate the origin to the left terminus by the change of variable $x = x' - x_0'$, and designate the left and right truncation points in standard units of the population (complete distribution) as ξ' and ξ'', respectively. We can write the probability density function for this case as

$$f(x) = \frac{1}{(I_0' - I_0'')\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\xi' + x/\sigma)^2}, \qquad 0 \leq x \leq R, \tag{1}$$

where

$$I_0' = \frac{1}{\sqrt{2\pi}} \int_{\xi'}^{\infty} e^{-t^2/2}\, dt, \qquad I_0'' = \frac{1}{\sqrt{2\pi}} \int_{\xi''}^{\infty} e^{-t^2/2}\, dt, \tag{2}$$

and

$$\mu = x_0' - \sigma\xi'. \tag{3}$$

Thus $(I_0' - I_0'')$ is the area under the normal curve between ordinates erected at ξ' and ξ'' respectively. Moreover $(I_0' - I_0'') = P(x_0' \leq x' \leq x_0' + R)$. The likelihood function for such a sample is

$$P(x_1, x_2, \cdots, x_{n_0}) = \left(\frac{1}{(I_0' - I_0'')\sigma\sqrt{2\pi}}\right)^{n_0} e^{-\frac{1}{2}\sum_1^{n_0}(\xi' + x_i/\sigma)^2}. \tag{4}$$

Since R is the truncated range, and since ξ' and ξ'' are in standard units, we have

$$\xi'' = \xi' + R/\sigma. \tag{5}$$

It should be understood that ξ' is considered throughout this paper, as the independent parameter of location. The mean, μ, cf. (3), is a linear function of ξ'.

In the derivations which follow, we employ the Fisher I_n functions, where $I_0(\xi)$ is defined by (2) and

$$I_n(\xi) = \int_{\xi}^{\infty} I_{n-1}(t)\, dt, \tag{6}$$

and hence

$$\frac{dI_n}{d\xi} = -I_{n-1}.$$

These functions satisfy the recurrence formula

$$(n + 1)I_{n+1} + \xi I_n - I_{n-1} = 0, \qquad n \geq -1. \tag{7}$$

$I_n(\xi)$ is ordinarily abbreviated to I_n in this paper. Where no confusion seems likely to occur, similar abbreviations are used for other functions of ξ.

We now obtain certain relations for use in subsequent derivations. Equations (2), (5), and (6) enable us to write

$$(8) \quad \frac{\partial I_0'}{\partial \xi'} = -I_{-1}' = -\varphi(\xi'), \quad \frac{\partial I_0''}{\partial \xi'} = -I_{-1}'' = -\varphi(\xi''), \quad \frac{\partial I_0''}{\partial \sigma} = -I_{-1}'' \frac{\partial \xi''}{\partial \sigma},$$

where $\varphi(\xi)$ is the ordinate of the normal frequency curve, i.e., $\varphi(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2}$. Ordinarily we abbreviate $\varphi(\xi')$ to φ' and $\varphi(\xi'')$ to φ''. On differentiating (5) we have

$$(9) \qquad \frac{\partial \xi''}{\partial \sigma} = -\frac{R}{\sigma^2}$$

and hence from (8)

$$\frac{\partial I_0''}{\partial \sigma} = \varphi'' \frac{R}{\sigma^2}.$$

Taking logarithms of (4), differentiating with the aid of (8) and (9), and equating to zero, we obtain the maximum likelihood estimating equations

$$(10) \quad \begin{aligned} \frac{\partial L}{\partial \xi'} &= \frac{n_0(\varphi' - \varphi'')}{I_0' - I_0''} - \sum_1^{n_0} \left(\xi' + \frac{x_i}{\sigma}\right) = 0, \\ \frac{\partial L}{\partial \sigma} &= \left(\frac{n_0 \varphi''}{I_0' - I_0''}\right) \frac{R}{\sigma^2} - \frac{n_0}{\sigma} + \frac{1}{\sigma^2} \sum_1^{n_0} \left\{x_i \left(\xi' + \frac{x_i}{\sigma}\right)\right\} = 0. \end{aligned}$$

If we define

$$(11) \qquad Z_1 = \frac{\varphi'}{I_0' - I_0''}, \qquad Z_2 = \frac{\varphi''}{I_0' - I_0''},$$

and substitute these values in (10), the estimating equations become

$$(12) \quad \begin{aligned} \sigma[Z_1 - Z_2 - \xi'] - \nu_1 &= 0, \\ \sigma^2[1 - \xi'(Z_1 - Z_2 - \xi') - Z_2 R/\sigma] - \nu_2 &= 0, \end{aligned}$$

where ν_1 and ν_2 are the first and second sample moments referred to the left terminus; i.e., $\nu_k = \sum_1^{n_0} x_i^k / n_0$.

To obtain the required estimates $\hat{\sigma}$ and $\hat{\xi}'$, it is necessary to solve the two equations of (12) simultaneously. As illustrated in Section 7, this can be accomplished without too much difficulty with the aid of the normal curve tables by using a modified Newton-Raphson method for solving two equations in two unknowns. This method is described in greater detail by Whittaker and Robinson [8]. Note that Z_1 and Z_2, cf. (11), involve only the normal curve ordinates φ' and φ'' and the areas I_0' and I_0''. Consequently they can be evaluated for any

desired values of ξ' and σ from standard tables of the normal frequency function. To determine $\hat{\mu}$, substitute $\hat{\sigma}$ and $\hat{\xi}'$ in (3).

Throught this paper, we designate the maximum likelihood estimates as $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}'$ respectively, whereas corresponding population parameters are designated as μ, σ, and ξ'

4. Case II. Number of unmeasured observations in each 'tail' known. Let the truncation points, the origin of reference, and the number of measured observations be designated as for Case I. If we let n_1 and n_2 be the number of unmeasured observations in the left and right 'tails' respectively, the likelihood function for a sample of this type is

$$(13)\quad P(x_1, x_2, \cdots, x_{n_1+n_0+n_2}) = K(1 - I_0')^{n_1}\cdot\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n_0} e^{-\frac{1}{2}\sum_1^{n_0}(\xi'+x_i/\sigma)^2}\cdot(I_0'')^{n_2},$$

where K is a constant.

We take the logarithms of (13), differentiate with the help of (8) and (9), and equate to zero to obtain the maximum likelihood estimating equations

$$(14)\quad \begin{aligned} \frac{\partial L}{\partial \xi'} &= n_1\frac{\varphi'}{1 - I_0'} - n_2\frac{\varphi''}{I_0''} - \sum_1^{n_0}\left(\xi' + \frac{x_i}{\sigma}\right) = 0, \\ \frac{\partial L}{\partial \sigma} &= n_2\frac{\varphi''}{I_0''} - \frac{n_0}{\sigma} + \frac{1}{\sigma^2}\sum_1^{n_0}\left\{x_i\left(\xi' + \frac{x_i}{\sigma}\right)\right\} = 0. \end{aligned}$$

Let

$$(15)\qquad Y_1 = \frac{n_1}{n_0}\frac{\varphi'}{(1 - I_0')}, \qquad Y_2 = \frac{n_2}{n_0}\frac{\varphi''}{I_0''},$$

and (14) can be written as

$$(16)\quad \begin{aligned} \sigma[Y_1 - Y_2 - \xi'] - \nu_1 &= 0, \\ \sigma^2[1 - \xi'(Y_1 - Y_2 - \xi') - Y_2R/\sigma] - \nu_2 &= 0, \end{aligned}$$

where ν_1 and ν_2 are again the first and second sample moments referred to the left terminus. The estimating equations (16) correspond to equations (12) given for Case I, and the manner of solution is the same for both cases. Y_1 and Y_2 for a given sample are functions of ξ' and σ only. They can be evaluated for any desired values of these variables from ordinary normal curve tables. As in Case I, the mean is estimated from (3).

5. Case III. Total number of unmeasured observations known, but not the number in each tail. Again, let the truncation points, the origin of reference, and the number of measured observations be designated as in the two previous cases. Let N be the total sample size and hence $N - n_0$ the combined number of

unmeasured observations in both tails. In the notation of Case II, $N - n_0 = n_1 + n_2$. The likelihood function for a sample of this type is

$$(17)\quad P(x_1, x_2, \cdots, x_N) = K(1 - I_0' + I_0'')^{N-n_0}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n_0} e^{-\frac{1}{2}\sum_1^{n_0}(\xi'+x_i/\sigma)^2}.$$

Taking logarithms of (17), differentiating with the assistance of (8) and (9) and equating to zero, we obtain the maximum likelihood estimating equations

$$(18)\quad \begin{aligned} \frac{\partial L}{\partial \xi'} &= (N - n_0)\left(\frac{\varphi' - \varphi''}{1 - I_0' + I_0''}\right) - \sum_1^{n_0}\left(\xi' + \frac{x_i}{\sigma}\right) = 0, \\ \frac{\partial L}{\partial \sigma} &= (N - n_0)\left(\frac{\varphi''}{1 - I_0' + I_0''}\right)\frac{R}{\sigma^2} - \frac{n_0}{\sigma} + \frac{1}{\sigma^2}\sum_1^{n_0}\left\{x_i\left(\xi' + \frac{x_i}{\sigma}\right)\right\} = 0. \end{aligned}$$

In this instance, let

$$(19)\quad Q_1 = \left(\frac{N - n_0}{n_0}\right)\frac{\varphi'}{1 - I_0' + I_0''}, \qquad Q_2 = \left(\frac{N - n_0}{n_0}\right)\frac{\varphi''}{1 - I_0' + I_0''},$$

and (18) can be written as

$$(20)\quad \begin{aligned} \sigma[Q_1 - Q_2 - \xi'] - \nu_1 &= 0, \\ \sigma^2[1 - \xi'(Q_1 - Q_2 - \xi') - Q_2R/\sigma] - \nu_2 &= 0. \end{aligned}$$

It will be recognized that equations (20) correspond to (12) and (16) for Cases I and II respectively. Since the manner of solving the estimating equations is identical in all three cases, it will not be discussed further here For any given sample, Q_1 and Q_2 are functions of ξ' and σ only, and they can be evaluated for any desired values of these arguments from standard normal curve tables. In this case also, the mean is estimated from equation (3).

6. First approximations.

CASE I. In this case, the following relations will usually provide satisfactory first approximations for estimating σ and ξ':

$$(21)\quad \sigma_1 = s_x, \qquad \xi_1' = -\nu_1/s_x,$$

where s_x^2 is the sample variance, i.e, $s_x^2 = (\nu_2 - \nu_1^2)$. It should be remarked that the only penalty involved in beginning with a poor first approximation is to increase slightly the number of steps necessary before arriving at a satisfactory final approximation by the method of Section 7.

CASE II. Since n_1 and n_2 are known in this case, it is more expedient to read first approximations to ξ' and ξ'' directly from standard tables of normal curve areas where we set

$$(22)\quad \frac{n_1}{n_1 + n_0 + n_2} = 1 - I_0' = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\xi'} e^{-t^2/2}\,dt,$$

and

$$\frac{n_2}{n_1 + n_0 + n_2} = I_0'' = \frac{1}{\sqrt{2\pi}} \int_{\xi''}^{\infty} e^{-t^2/2}\, dt. \tag{23}$$

With ξ' and ξ'' determined from (22) and (23), we obtain a first approximation for estimating σ, from equation (5), which we now write as

$$\sigma_1 = R/(\xi_1'' - \xi_1'). \tag{24}$$

Case III. For a first approximation in this case, it will usually be satisfactory, in the absence of contrary information, to assume that the unmeasured observations are divided equally between the two tails, and then proceed as in Case II.

7. Numerical examples. As previously mentioned, a modified Newton-Raphson method for solving two equations in two unknowns is satisfactory in each of the three cases considered, for solving the estimating equations to obtain $\hat{\sigma}$ and ξ' in practical applications. A random sample from a normal population with $\mu = 0$, and $\sigma = 1$, selected from Mahalanobis's tables [9] will serve to illustrate the solution in each case.

Case I. For the sample selected, $n_0 = 32$; $\nu_1 = 1.244625$; $\nu_2 = 2.105275$; $x_0' = -1.000000$; and $R = 2.750000$. The estimating equations to be solved simultaneously for ξ' and $\hat{\sigma}$ are thus

$$\sigma[Z_1 - Z_2 - \xi'] - 1.244625 = 0,$$

$$\sigma^2[1 - \xi'(Z_1 - Z_2 - \xi') - 2.750000\, Z_2/\sigma] - 2.105275 = 0.$$

For first approximations, we employ (21) to obtain; $\sigma_1 = s_x = 0.75$; and $\xi_1' = -1.244625/0.75 = -1.66$. Beginning with these approximations, we subsequently obtain the results displayed in Table 1.

TABLE 1

Solution of estimating equations in Case I

σ	ξ' from ν_1	ξ' from ν_2	Difference
1.536313	−0.5389	−0.5387	−0.0002
1.527778	−0.5455	−0.5460	+0.0005

Interpolating in this table, we obtain $\hat{\sigma} = 1.534$ and $\hat{\xi}' = -0.541$. On substituting these values in (3) we obtain $\hat{\mu} = -0.170$. Even though the first approximations in this instance proved to be considerably in error, no appreciable increase was experienced in the number of steps necessary to arrive at the final values given.

Case II. Solution of estimating equations (16) for this case can also be illustrated with the same sample which was used in Case I. In this instance, however,

we have the additional information; $n_1 = 7$ and $n_2 = 1$. The equations to be solved are:

$$\sigma[Y_1 - Y_2 - \xi'] - 1.244625 = 0,$$

$$\sigma^2[1 - \xi'(Y_1 - Y_2 - \xi') - 2.750000\, Y_2/\sigma] - 2.105275 = 0.$$

From (22), (23) and (24) we obtain the first approximations: $\xi_1' = -0.935$; $\xi_1'' = 1.960$; and hence $\sigma_1 = 0.950$. Beginning with these values, we proceed as in Case I, and after several trials obtain the results displayed in Table 2.

TABLE 2

Solution of estimating equations in Case II

σ	ξ' from ν_1	ξ' from ν_1	Difference
1.041667	−0 9381	−0.9360	−0 0021
1.000000	−0 9820	−1.0094	+0.0274

Interpolating, we have $\hat{\sigma} = 1.039$ and $\hat{\xi}' = -0.941$. From (3) we then obtain. $\hat{\mu} = -0.022$.

CASE III. Again we use the same sample that was employed to illustrate Cases I and II. In this instance, however, we assume that the only information available about the unmeasured observations is that their total number is 8. In the notation of Section 5, we have $N = 40$, $n_0 = 32$, and hence $N - n_0 = 8$. The estimating equations in this situation are

$$\sigma[Q_1 - Q_2 - \xi'] - 1.244625 = 0,$$

$$\sigma^2[1 - \xi'(Q_1 - Q_2 - \xi') - 2.750000\, Q_2/\sigma] - 2\ 105275 = 0.$$

Under the assumption that 4 unmeasured observations are in each 'tail', equations (22), (23) and (24) give first approximations: $\xi_1' = -1.28$; $\xi_1'' = 1.28$; and hence $\sigma_1 = 1.074$. Starting with these values and proceeding as in the two previous cases, we obtain the results displayed in Table 3.

TABLE 3

Solution of estimating equations in Case III

σ	ξ' from ν_1	ξ' from ν_2	Difference
1.000000	−1.0794	−1.2091	+0 1297
1.100000	−1.0118	−0.9739	−0 0379

By interpolation, we have $\hat{\sigma} = 1.077$ and $\hat{}' = -1.027$. From equation (3), we then compute $\hat{\mu} = 0.106$.

8. Precision of estimates. To determine asymptotic variances of $\hat{\sigma}$ and $\hat{}'$, we construct the variance-covariance matrices. This requires that we obtain the

second partial derivatives of logarithms of the likelihood function in each of the three cases considered. Results stated in (8) and (9) are involved in these derivatives.

Case I. The second partial derivatives in this case are

$$(25)\quad \frac{\partial^2 L}{\partial \xi'^2} = n_0 f_1(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \xi' \partial \sigma} = \frac{n_0}{\sigma} f_2(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \sigma^2} = \frac{n_0}{\sigma^2} f_3(\xi', \xi'');$$

where

$$(26)\quad \begin{aligned} f_1(\xi', \xi'') &= -[1 + \xi' Z_1 - \xi'' Z_2 - (Z_1 - Z_2)^2], \\ f_2(\xi', \xi'') &= \left\{\frac{R}{\sigma} Z_2[(Z_1 - Z_2) - \xi''] + [Z_1 - Z_2 - \xi']\right\}, \\ f_3(\xi', \xi'') &= \left\{\left(\frac{R}{\sigma}\right)^2 Z_2(Z_2 + \xi'') - \left[2 - \xi'(Z_1 - Z_2 - \xi') - Z_2 \frac{R}{\sigma}\right]\right\}. \end{aligned}$$

Subsequently we obtain

$$(27)\quad V(\hat{\sigma}) = \frac{\sigma^2}{n_0}\left[\frac{-f_1}{f_1 f_3 - f_2^2}\right], \qquad V(\hat{\xi}') = \frac{1}{n_0}\left[\frac{-f_3}{f_1 f_3 - f_2^2}\right], \qquad r_{\hat{\sigma}, \hat{\xi}'} = \frac{f_2}{\sqrt{f_1 f_3}}.$$

Case II. In this case the second partial derivatives are

$$(28)\quad \frac{\partial^2 L}{\partial \xi'^2} = n_0\, g_1(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \xi' \partial \sigma} = \frac{n_0}{\sigma} g_2(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \sigma^2} = \frac{n_0}{\sigma^2} g_3(\xi', \xi''),$$

where

$$(29)\quad \begin{aligned} g_1(\xi', \xi'') &= -\left[1 + \xi' Y_1 - \xi'' Y_2 + \frac{n_0}{n_1} Y_1^2 + \frac{n_0}{n_2} Y_2^2\right], \\ g_2(\xi', \xi'') &= \left\{\frac{R}{\sigma} Y_2\left[\frac{n_0}{n_2} Y_2 - \xi''\right] + [Y_1 - Y_2 - \xi']\right\}. \\ g_3(\xi', \xi'') &= \left\{\left(\frac{R}{\sigma}\right)^2 Y_2\left(\xi'' - \frac{n_0}{n_2} Y_2\right) - [2 - \xi'(Y_1 - Y_2 - \xi') - Y_2 R/\sigma]\right\}. \end{aligned}$$

Finally we can write

$$(30)\quad V(\hat{\sigma}) = \frac{\sigma^2}{n_0}\left[\frac{-g_1}{g_1 g_3 - g_2^2}\right], \qquad V(\hat{\xi}') = \frac{1}{n_0}\left[\frac{-g_3}{g_1 g_3 - g_2^2}\right], \qquad r_{\hat{\sigma}, \hat{\xi}'} = \frac{g_2}{\sqrt{g_1 g_3}}.$$

Case III. This time, the second partial derivatives are

$$(31)\quad \frac{\partial^2 L}{\partial \xi'^2} = n_0 h_1(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \xi' \partial \sigma} = \frac{n_0}{\sigma} h_2(\xi', \xi''), \qquad \frac{\partial^2 L}{\partial \sigma^2} = \frac{n_0}{\sigma^2} h_3(\xi', \xi''),$$

where

$$
(32)\quad
\begin{aligned}
h_1(\xi', \xi'') &= -\left[1 + \xi' Q_1 - \xi'' Q_2 + \frac{n_0}{N - n_0}(Q_1 - Q_2)^2\right],\\
h_2(\xi', \xi'') &= \left\{\frac{R}{\sigma} Q_2 \left[\left(\frac{n_0}{N - n_0}\right)(Q_2 - Q_1) - \xi''\right] + [Q_1 - Q_2 - \xi']\right\},\\
h_3(\xi', \xi'') &= \left\{\left(\frac{R}{\sigma}\right)^2 Q_2 \left(\xi'' - \frac{n_0}{N - n_0} Q_2\right)\right.\\
&\qquad \left. - \left[2 - \xi'(Q_1 - Q_2 - \xi') - Q_2 \frac{R}{\sigma}\right]\right\}.
\end{aligned}
$$

Accordingly we obtain

$$
(33)\quad V(\hat{\sigma}) = \frac{\sigma^2}{n_0}\left[\frac{-h_1}{h_1 h_3 - h_2^2}\right], \quad V(\hat{\xi}') - \frac{1}{n_0}\left[\frac{-h_3}{h_1 h_3 - h_2^2}\right], \quad r_{\hat{\sigma},\hat{\xi}'} = \frac{h_2}{\sqrt{h_1 h_3}}.
$$

Note that variances of the estimates for each case considered, can be computed for given values of ξ' and σ from standard normal tables of areas and ordinates

9. Singly truncated samples. If only the left 'tail' is missing from the samples thus far considered, then $\xi'' = \infty$, $n_2 = 0$, $\varphi'' = 0$, $I_0'' = 0$, and hence Z_2, Y_2, and Q_2 each equal zero. Upon substituting these values in (12), (16), and (20) respectively, estimating equations applicable to singly truncated samples are obtained as special cases of the estimating equations for doubly truncated samples. Of course Cases II and III become identical when samples are singly truncated When $Y_2 = Q_2 = 0$, then $Y_1 = Q_1$, cf. (15) and (19)

CASE I. With $Z_2 = 0$, the estimating equations (12) become

$$
(34)\quad
\begin{aligned}
\sigma[Z_1 - \xi'] &= \nu_1,\\
\sigma^2[1 - \xi'(Z_1 - \xi')] &= \nu_2.
\end{aligned}
$$

Eliminating σ between these two equations we have

$$
(35)\quad \frac{\nu_2}{\nu_1^2} = \frac{1}{Z_1 - \xi'}\left(\frac{1}{Z_1 - \xi'} - \xi'\right),
$$

which is recognized as the Pearson-Lee-Fisher equation in a form which was previously given by the author [5].

CASE II. With $Y_2 = 0$, the estimating equations (16) become

$$
(36)\quad
\begin{aligned}
\sigma[Y_1 - \xi'] &= \nu_1\\
\sigma^2[1 - \xi'(Y_1 - \xi')] &= \nu_2.
\end{aligned}
$$

Eliminating σ between the above equations, we obtain

$$
(37)\quad \frac{\nu_2}{\nu_1^2} = \frac{1}{Y_1 - \xi'}\left(\frac{1}{Y_1 - \xi'} - \xi'\right),
$$

which is in a form completely analogous to (35). Furthermore, this equation can be solved for $\hat{\xi}'$ in the same manner as (35), cf. [5]. Since σ can be eliminated between estimating equations in singly truncated cases, but not in doubly

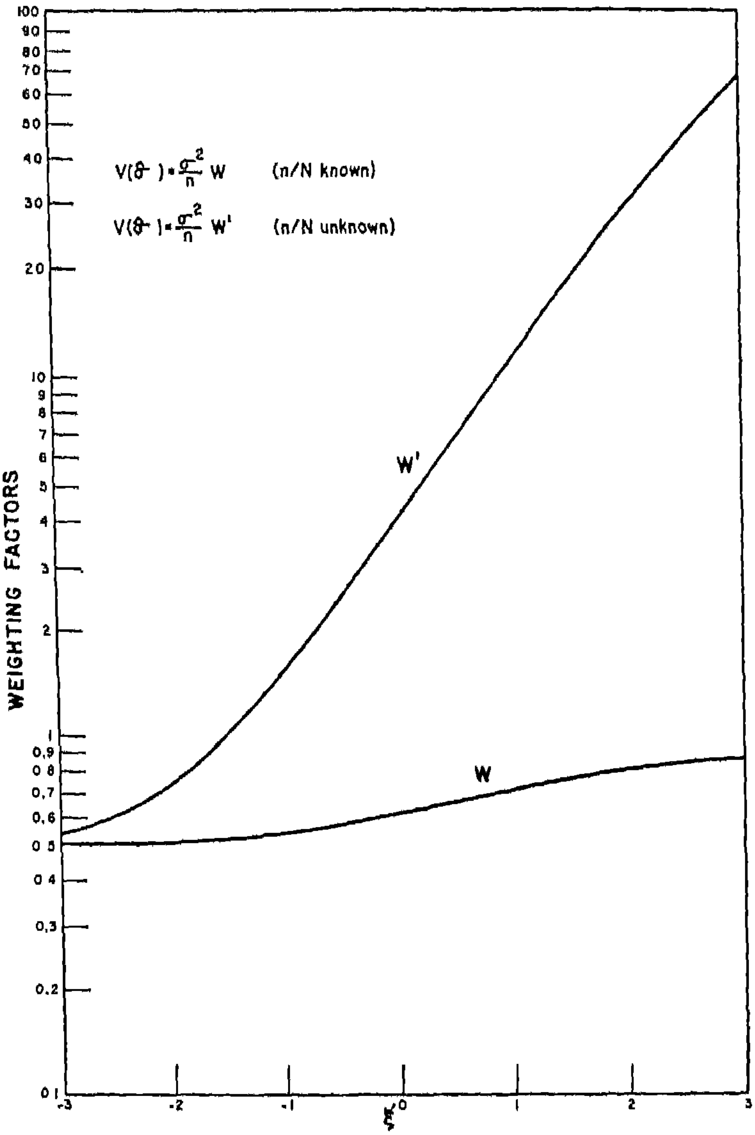


FIG. 1. Weighting factors for use in determining the variance of $\hat{\sigma}$.

truncated cases, the numerical computations are much simpler and less laborious for singly truncated samples.

If the right rather than the left tail is missing from singly truncated samples,

pplicable estimating equations can be obtained from (12) and (16) by translating he origin to the terminus on the right and setting Z_1 and Y_1 equal to zero ather than Z_2 and Y_2.

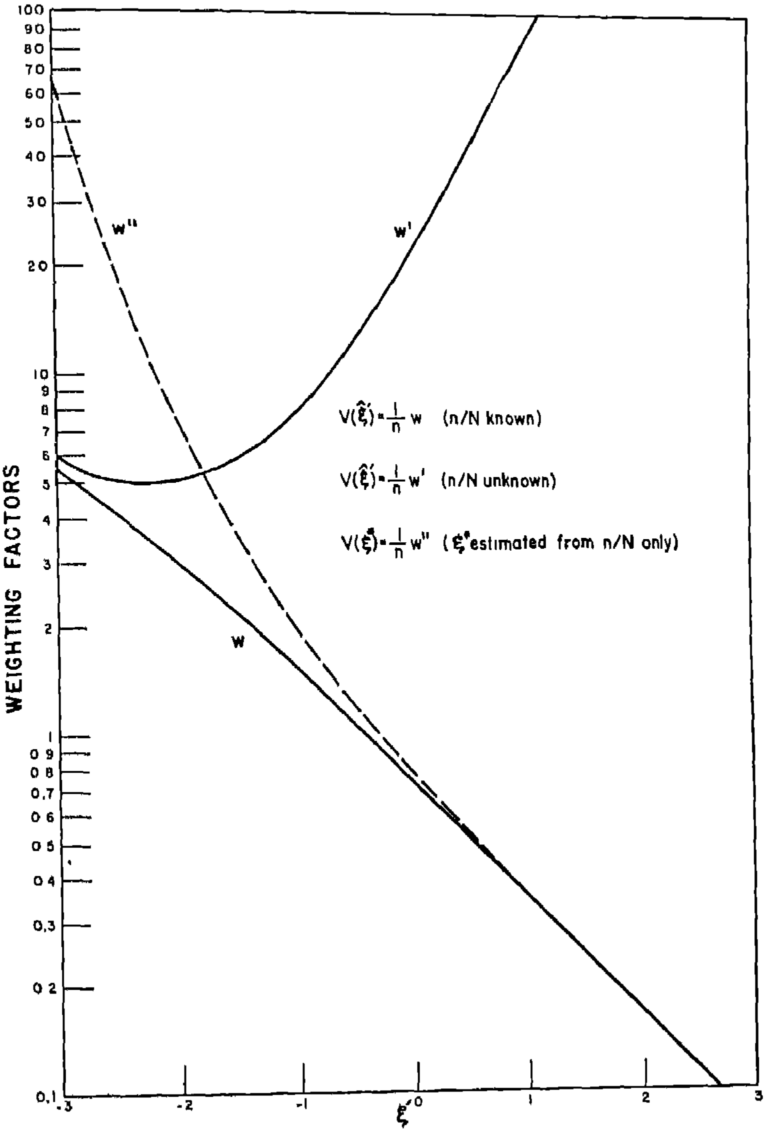


FIG. 2. Weighting factors for use in determining the variances of $\hat{\xi}'$ and ξ^*

The variance formulas (25) and (28) likewise assume more simple forms with singly truncated samples. Substitute $Z_2 = 0$ in (25) and the variance formulas applicable with singly truncated samples when the number of unmeasured

observations is unknown, become identical in form with those previously given by the writer in [5]. When the number of unmeasured observations in a singly truncated sample is known, the applicable variance formulas (28), on setting $Y_2 = 0$, become

$$V(\hat{\sigma}) = \frac{\sigma^2}{n} W(\xi') \quad \text{and} \quad V(\hat{\xi}') = \frac{1}{n} w(\xi'), \tag{38}$$

where W and w may be regarded as weighting functions defined by

$$W(\xi') = \frac{1 + Y_1(Y_1 n_0/n_1 + \xi')}{[2 - \xi'(Y_1 - \xi')][1 + Y_1(Y_1 n_0/n_1 + \xi')] - [Y_1 - \xi']^2} \tag{39}$$

and

$$w(\xi') = \frac{2 - \xi'(Y_1 - \xi')}{[2 - \xi'(Y_1 - \xi')][1 + Y_1(Y_1 n_0/n_1 + \xi')] - [Y_1 - \xi']^2}. \tag{40}$$

Similarly, the correlation between sampling errors of $\hat{\sigma}$ and $\hat{\xi}'$ in this case becomes

$$r_{\hat{\sigma}\cdot\hat{\xi}'} = \frac{Y_1 - \xi'}{\sqrt{[2 - \xi'(Y_1 - \xi')][1 + Y_1(Y_1 n_0/n_1 + \xi')]}}. \tag{41}$$

A comparison of the variances (38), with those applicable when the number of unmeasured observations is unknown, serves to indicate the extent to which information contained in a singly truncated sample is increased by adding knowledge of the number of unmeasured observations. To facilitate such comparisons, W, w, and corresponding functions W' and w' applicable when the number of unmeasured observations is unknown, are displayed graphically in Figures 1 and 2. In computing the plotted values of W and w, the ratio n/N in (39) and (40) was replaced by I_0. This ratio is, of course, an estimate of I_0, and for n and N sufficiently large, the substitution is amply justified. Equations for W' and w' can be found in [5]. For further comparisons, a graph of w'' applicable in determining the variance $V(\xi^*)$, where ξ^* is estimated from n/N alone is also included in Figure 2. This latter function is defined as

$$w''(\xi^*) = \frac{I_0^2(1 - I_0)}{\varphi^2}. \tag{42}$$

It follows from the well known formula for the variance of ξ^*:

$$V(\xi^*) = \frac{1}{N}\left\{\frac{I_0(1 - I_0)}{\varphi_2}\right\} = \frac{1}{n}\left\{\frac{I_0^2(1 - I_0)}{\varphi_2}\right\}. \tag{43}$$

An examination of Figures 1 and 2 discloses that except when the omitted portion of the distribution is small ($\xi' < -3$), the variances of the estimates of σ and ξ' based on singly truncated normal samples are substantially less when the number of unmeasured observations is known than when this information is lacking.

REFERENCES

[1] K Pearson and A Lee, "On the generalized probable error in multiple normal correlation", *Biometrika*, Vol. 6 (1908), pp 59–68

[2] A Lee, "Table of Gaussian 'tail' functions when the 'tail' is larger than the body", *Biometrika*, Vol 10 (1915), pp 208–215

[3] R. A. Fisher, "Properties and applications of H_h functions", *Mathematical Tables*, Vol. 1, pp xxvi–xxxv, British Association for the Advancement of Science, 1931

[4] A. Hald, "Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point", *Skandinavisk Aktuarietidskrift*, Vol. 32 (1949), pp 119–134.

[5] A C Cohen, Jr., "On estimating the mean and standard deviation of truncated normal distributions", *Jour Am Stat. Assn*, Vol. 44 (1949), pp 518–525.

[6] W. L Stevens, "The truncated normal distribution", appendix to "The Calculation of the Time-Mortality Curve" by C I. Bliss, *Annals of Applied Biology*, Vol 24 (1937), pp. 815–852.

[7] W G. Cochran, "Use of IBM equipment in an investigation of the truncated normal problem", *Proc Research Forum*, International Business Machines Corp , 1946, pp. 40–43.

[8] E T. Whittaker and G Robinson, *The Calculus of Observations*, Second Ed , Blackie and Son, Ltd., London and Glasgow, 1929, pp. 88–91.

[9] P. C Mahalanobis, "Tables of random samples from a normal population", *Sankhyā*, Vol 1 (1934), pp. 289–328.

THE ASYMPTOTIC PROPERTIES OF ESTIMATES OF THE PARAMETERS OF A SINGLE EQUATION IN A COMPLETE SYSTEM OF STOCHASTIC EQUATIONS[1, 2]

By T. W. Anderson[3] and Herman Rubin[4]

Columbia University and Institute for Advanced Study

1. Summary. In a previous paper [2] the authors have given a method for estimating the coefficients of a single equation in a complete system of linear stochastic equations. In the present paper the consistency of the estimates and the asymptotic distributions of the estimates and the test criteria are studied under conditions more general than those used in the derivation of these estimates and criteria. The point estimates, which can be obtained as maximum likelihood estimates under certain assumptions including that of normality of disturbances, are consistent even if the disturbances are not normally distributed and (a) some predetermined variables are neglected (Theorem 1) or (b) the single equation is in a non-linear system with certain properties (Theorem 2).

Under certain general conditions (normality of the disturbances not being required) the estimates are asymptotically normally distributed (Theorems 3 and 4). The asymptotic covariance matrix is given for several cases. The criteria derived in [2] for testing the hypothesis of over-identification have, asymptotically, χ^2-distributions (Theorem 5). The exact confidence regions developed in [2] for the case that all predetermined variables are exogenous (that is, that the difference equations are of zero order) are shown to be consistent and to hold asymptotically even when this assumption is not true (Theorem 6).

2. Introduction. The complete system of linear stochastic equations considered by the authors in [2] was written

$$\mathrm{B}_{yy} y_t' + \Gamma_{yz} z_t' = \epsilon_t', \tag{2.1}$$

where y_t is a row vector of G jointly dependent variables at "time" t, z_t is a row vector of K variables predetermined at t, and ϵ_t is a row vector of "disturbances," and B_{yy} and Γ_{yz} are matrices. If B_{yy} is non-singular the distribution of ϵ_t induces the distribution of y_t given z_t.

One component equation of (2.1) was given special treatment. Let β be

[1] This paper will be included in Cowles Commission Papers, New Series, No. 36.

[2] The results of this paper were presented to meetings of the Institute of Mathematical Statistics at Washington, D. C., April 12, 1946 (Washington Chapter) and at Ithaca, New York, August 23, 1946. Most of the research was done at the Cowles Commission for Research in Economics; the authors are indebted to the members of the Cowles Commission staff for many helpful discussions

[3] Fellow of the John Simon Guggenheim Memorial Foundation; Research Consultant of the Cowles Commission for Research in Economics

[4] National Research Fellow, Research Consultant of the Cowles Commission for Research in Economics

composed of the coefficients of the coordinates of y_t which are not assumed zero in the specified equation, and let x_t be composed of the corresponding components of y_t; similarly let γ be composed of the coefficients of the coordinates of z_t which are not assumed zero, and u_t the corresponding components of z_t; and let ζ_t be the component of ϵ_t associated with the specified equation. Then the single equation is

$$\beta x_t' + \gamma u_t' = \zeta_t. \tag{2.2}$$

Suppose we have a set of observations $x_t, z_t, t = 1, \cdots, T$. For sets of any two vectors a_t and b_t, let the second-order moment matrix be

$$M_{ab} = \frac{1}{T}\sum_{t=1}^{T} a_t' b_t. \tag{2.3}$$

Let s_t be some linear transform of v_t, the set of coordinates of z_t not contained in u_t, chosen so $M_{su} = 0$. Defining

$$W_{xx} = M_{xx} - M_{xz}M_{zz}^{-1}M_{zx}, \tag{2.4}$$

and assuming ϵ_t normally distributed with mean 0, covariance matrix Σ, and independently of $\epsilon_{t'}(t \neq t')$, we find $\hat\beta$, the maximum likelihood estimate of β, to be proportional to a vector defined by

$$(M_{xs}M_{ss}^{-1}M_{sx} - \nu W_{xx})b' = 0, \tag{2.5}$$

taking ν as the smallest root of

$$| M_{xs}M_{ss}^{-1}M_{sx} - \nu W_{xx} | = 0. \tag{2.6}$$

The vector is normalized by

$$\hat\beta\hat\Phi_{xx}\hat\beta' = 1, \tag{2.7}$$

where Φ_{xx} may be a function of the estimates of other parameters. The estimate of γ is $\hat\gamma = -\hat\beta M_{xu}M_{uu}^{-1}$ [2; Theorem 1]. These estimates were derived under the following explicit Assumptions A, B, C, and D:

Assumption A. *The selected structural equation* (2.2) *is one equation of a complete linear system of stochastic equations. It is identified by the fact that if H is the number of coordinates in x_t, there are at least $H - 1$ coordinates in v_t, the vector of predetermined variables in the system, but missing in* (2.2).

Assumption B. *At time t all of the coordinates of $z_t = (u_t, v_t)$ are given.*

Assumption C. *The coordinates of z_t are given functions of exogenous variables and of coordinates of $y_{t-1}, y_{t-2}, \cdots$. If coordinates of $y_0, y_{-1}, \cdots$ are involved in z_t, they will be considered as given numbers The moment matrix M_{zz} is non-singular with probability one.*

Assumption D. *The disturbance vectors ϵ_t are distributed serially independently and normally with mean zero and covariance matrix Σ_{xx}.*

Under these assumptions it is found that $(1 + \nu)^{-\frac{1}{2}T}$ is the likelihood ratio

criterion for testing the hypothesis that the number of components of z_t assumed to have zero coefficients is so great.

If there are no lagged endogenous variables in z_t, we can find confidence regions for β and for β and γ simultaneously as well as an approximate test for the above hypothesis. The assumptions used for these results are A, B, and

ASSUMPTION E. *All the coordinates of* $z_t = (u_t, v_t)$ *are exogenous. The moment matrix* M_{zz} *is non-singular. The disturbances of the selected equation are distributed independently and normally with mean zero and variance* σ^2.

Assumptions A and B are used in this paper and a number in addition, which will be lettered similarly. It is to be emphasized that the various assumptions are used alternatively, never all at once; in fact many assumptions are mutually exclusive.

3. Consistency of the estimates. The estimates $\hat{\beta}$ and $\hat{\gamma}$ are consistent not only in the case for which they are maximum likelihood estimates, but also in cases in which the disturbances are not normally or even identically distributed. Moreover, for consistency of the estimates it is not necessary that the investigator know all of the components of v_t or use them. Another direction in which the assumptions may be relaxed is to permit the other equations in the system to be non-linear.

3.1. *The linear case.* This case is characterized by Assumption A. We need also to assume:

ASSUMPTION F. M_{zz} *converges to a fixed non-singular limit* R *in probability.*

Let u_t consist of the part of z_t that enters the selected structural equation (22). The remainder of the components of z_t are divided into two groups as to whether they are known or not. Let c_t be a linear transform of the known components not entering the specified equation such that

$$\operatorname*{plim}_{t\to\infty} M_{uc} = 0, \tag{3.1}$$

and let r_t be a linear transform of the components of z_t not known such that

$$\operatorname*{plim}_{t\to\infty} M_{ur} = 0, \tag{3.2}$$

$$\operatorname*{plim}_{t\to\infty} M_{cr} = 0. \tag{3.3}$$

The relevant part of the "reduced form," obtained from (2.1) by multiplication by B_{yy}^{-1} is

$$x_t' = \bar{\Pi}_{xu}u_t' + \Pi_{xc}c_t' + \Pi_{xr}r_t' + \delta_t'. \tag{3.4}$$

The matrix $(\Pi_{xc}\Pi_{xr})$ is Π_{xz} (defined in [2]) multiplied on the right by a non-singular matrix; hence, $\beta\Pi_{xc} = 0$, and similarly $\beta\bar{\Pi}_{xu} = \gamma$. We shall find it convenient to assume

ASSUMPTION G. Π_{xc} *has rank* $H - 1$.

This means that for T sufficiently large the probability is arbitrarily near 1 that (2.2) is identified.

However, these conditions still do not insure consistency. We need the asymptotic analogue of lack of correlation:

ASSUMPTION H.

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \delta_t' z_t = 0.$$

We do not need to require that the covariance matrices of δ_t are the same or even that they exist. We shall make an assumption about

$$W_{xx}^{*} = M_{xx} - (M_{xu} M_{xc}) \begin{pmatrix} M_{uu} & M_{uc} \\ M_{cu} & M_{cc} \end{pmatrix}^{-1} \begin{pmatrix} M_{ux} \\ M_{cx} \end{pmatrix}. \tag{3.5}$$

ASSUMPTION I. *The ratio of the largest to the smallest characteristic roots of W_{xx}^{*} is bounded in probability.*

This means that for a suitable constant K

$$\lim_{T\to\infty} P\left(\frac{l(W_{xx}^{*})}{s(W_{xx}^{*})} > K\right) = 0, \tag{3.6}$$

where $P(E)$ denotes the probability of event E and $s(A)$ and $l(A)$ are the smallest and largest roots of the matrix A, respectively

Assumptions F and H imply that $P_{xu} \to \Pi_{xu}$ and $P_{xc} \to \Pi_{xc}$ in probability, where $P_{xu} = M_{xu} M_{uu}^{-1}$ and P_{xc} is the part of

$$(M_{xu} M_{xc}) \begin{pmatrix} M_{uu} & M_{uc} \\ M_{cu} & M_{cc} \end{pmatrix}^{-1} \tag{3.7}$$

corresponding to the vector[5] c_t. The first assertion follows because $M_{xu} M_{uu}^{-1} = (\Pi_{xu} M_{uu} + \Pi_{xc} M_{xc} + \Pi_{xr} M_{ru} + M_{\delta u}) M_{uu}^{-1}$ and $M_{xc} \to 0$, $M_{ru} \to 0$, and $M_{\delta u} \to 0$ in probability by (3.1), (3.3) and Assumption H, the second assertion follows similarly. Since matrix multiplication is continuous, and the characteristic roots of a matrix are continuous functions of the matrix,[6]

$$\operatorname*{plim}_{T\to\infty} s[P_{xc} M_{**} P_{xc}'] = 0, \tag{3.8}$$

where $M_{**} = (M_{cc} - M_{cu} M_{uu}^{-1} M_{uc})$. This follows from the well-known theorem (a proof of which is given in [4]) that if a random vector X_T converges stochastically to X, then $f(X_T)$ converges stochastically to $f(X)$ if $f(y)$ is continuous at X.

We shall find the following lemmas convenient. The proofs are simple and have been given in [1].

LEMMA 1. *Let B be positive definite, A positive semi-definite. Then the smallest root ν of $|A - xB| = 0$ is less than or equal to $s(A)/s(B)$.*

[5] See Section 4 of [2].

[6] Because of the assertion above and Assumptions F and G only one characteristic root of the matrix approaches zero in probability.

LEMMA 2. *Each element of a positive definite matrix is less in absolute value than the largest characteristic root.*

Let ν be the smallest root of

$$(3.9) \qquad | P_{xc} M_{ss} P'_{xc} - \nu W^*_{xx} | = 0.$$

Then $\operatorname*{plim}_{T\to\infty} \nu W^*_{xx} = 0$. This statement follows from (3.8) and Lemmas 1 and 2.

Since 0 is a simple characteristic root of $\Pi_{xc} \operatorname*{plim}_{T\to\infty} M_{ss} \Pi'_{xc}$, it follows from (3.9) and the consistency of P_{xu} and P_{xc} that $\hat{\beta}$ approaches β apart from normalization. The following theorem results directly:

THEOREM 1. *Under Assumptions* A, F, G, H, *and* I, *and if* $\operatorname*{plim}_{T\to\infty} \beta\hat{\Phi}_{xx}\beta' = 1$,

$$(3.10) \qquad \operatorname*{plim}_{T\to\infty} \hat{\beta} = \beta,$$

$$(3.11) \qquad \operatorname*{plim}_{T\to\infty} \hat{\gamma} = \gamma,$$

where $\hat{\beta}$ and $\hat{\gamma}$ are calculated as if $r_t = 0$ and as if the remainder of A, B, C, *and* D *were satisfied.*[7]

3.2. *The non-linear case.* In this section we apply the estimates obtained in [2] to an equation of a complete system in which the remaining equations may be non-linear. We replace Assumption A by the following assumption:

ASSUMPTION J. *The selected structural equation* (2.2) *is one equation of a complete system of stochastic equations*:

$$(3.11) \qquad F_i(y_t, z_t) = \epsilon_{ti} \qquad (i = 1, \cdots, G).$$

Let us solve the complete system (3.11) for the components of y_t. We obtain

$$(3.12) \qquad y_{tj} = h_j(z_t, \epsilon_t).$$

Let u_t be the subvector of z_t occurring in the selected structural equation. Let c_t be a vector function of z_t such that $\operatorname*{plim}_{T\to\infty} M_{cu} = 0$. We may write (3.12) for those y's occurring in the selected structural equation as

$$(3.13) \qquad x'_t = \bar{\Pi}_{xu} u'_t + \Pi_{xc} c'_t + \varphi'(z_t, \epsilon_t),$$

where the components of $\varphi(z_t, \epsilon_t)$ are the residuals from the formal limiting regression of x_t on u_t and c_t. The proof of Theorem 1 can be used to prove the following:

THEOREM 2. *If Assumptions* F, G, H, I, *and* J *are satisfied with z_t replaced by (u_t, c_t) and δ_t replaced by $\varphi(z_t, \epsilon_t)$, and $r_t = 0$, and if* $\operatorname*{plim}_{T\to\infty} \beta\hat{\Phi}_{xx}\beta' = 1$, *then*

$$(3.14) \qquad \operatorname*{plim}_{T\to\infty} \hat{\beta} = \beta,$$

$$(3.15) \qquad \operatorname*{plim}_{T\to\infty} \hat{\gamma} = \gamma.$$

---

[7] This follows from the above statements because $\hat{\beta}$ and $\hat{\gamma}$ are (vector-valued) rational functions of M_{zz}, P_{xz}, W^*_{xx} and Φ_{xx} which approach limits in probability.

4. The asymptotic distribution of the estimates.

4.1. *The asymptotic distribution of P_{zz} and P_{zu}.* To obtain the asymptotic distribution of the estimates we need stronger assumptions. Throughout Sections 4.1 and 4.2 we use Assumptions A, B, F, H, I, and the following:

ASSUMPTION K. *The exogenous variables are bounded, the vector of disturbances of the complete system has mean zero, and is serially independent; for some $\lambda > 0$ and some M, $\mathcal{E}(|\delta_{ti}|^{4+\lambda}) < M$; the coordinates of z_t may be linear combinations of lagged endogenous variables. If the endogenous part of a coordinate is*

$$\sum_{\tau=1}^{\infty} \sum_{i=1}^{G} g_{\tau i} y_{t-\tau, i},$$

then

$$\sum_{\tau=1}^{\infty} \sum_{i=1}^{G} | g_{\tau i} | < \infty$$

and

$$\sum_{\tau=t}^{\infty} \sum_{i=1}^{G} g_{\tau i} y_{t-\tau, i}$$

is bounded.

ASSUMPTION L. *The matrix Φ_{xx} is known and constant.*

ASSUMPTION M. *For each i, j, k, l, $1 \leq i, j \leq H$, $1 \leq k, l \leq K$,*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathcal{E}(\delta_{ti}\delta_{tj} z_{tk} z_{tl}) = \kappa_{ijkl}$$

exists.

Let the components of M_{yy}, M_{yz}, M_{zz} be arranged as a vector $m(T)$ with mean value $\mu(T)$. It has been shown [3] that $\sqrt{T}(m(T) - \mu(T))$ is asymptotically distributed according to $N(0, \Sigma)$, the normal distribution with mean 0 and covariance matrix Σ composed of elements

$$\sigma_{ij} = \lim_{T\to\infty} \mathcal{E}(T[m_i(T) - \mu_i(T)]\,[m_j(T) - \mu_j(T)]).$$

In conjunction with this result we make repeated use of a special case of Theorem 6 of [4]:

Suppose $\sqrt{T}(x_{jT} - \xi_{jT})$ $(j = 1, \cdots, n)$ have the joint asymptotic distribution $N(0, \Psi)$ with ξ_{jT} being functions of T such that $\lim_{T\to\infty} \xi_{jT} = \xi_j$. Let $f_{kT}(z_1, \cdots, z_n)$ be random Borel-measurable functions of n real variables such that $\frac{\partial f_{kT}}{\partial z_j} = \alpha_{kjT}(z)$ exists with probability one for T sufficiently large and z in a fixed neighborhood of ξ, and suppose that there exist numbers α_{kj} such that for any $\epsilon > 0$, and $\lambda > 0$, $P(\sup_{(z-\xi_T)(z-\xi_T)' \leq (\lambda/T)} | \alpha_{kjT}(z) - \alpha_{kj} | > \epsilon)$ approaches zero. Then if $y_{kT} = f_{kT}(x_{1T}, \cdots, x_{nT})$ and $\eta_{kT} = f_{kT}(\xi_{1T}, \cdots, \xi_{nT})$, the random variables $\sqrt{T}(y_{kT} - \eta_{kT})$ have the joint asymptotic distribution $N(0, A\Psi A')$, where $A = (\alpha_{ij})$.

To obtain the asymptotic distributions we have only to verify that the assumptions of this statement are satisfied, and compute A, since the asymptotic distribution is characterized completely by $A\Psi A'$. We shall denote the element in the k-th row and l-th column of $A\Psi A'$ by $\sigma(f_k, f_l)$. We shall find it convenient to use the notation $df = Adx$; that is, the differential df is defined in terms of the limit matrix A.

Let

$$A = M_{bu}, \tag{4.1}$$

$$B = M_{bs}, \tag{4.2}$$

$$C = \operatorname*{plim}_{T\to\infty} M_{uu}, \tag{4.3}$$

$$E = \operatorname*{plim}_{T\to\infty} M_{ss}, \tag{4.4}$$

$$L = P_{xu}, \tag{4.5}$$

$$P = P_{xs} = M_{xs}M_{ss}^{-1}, \tag{4.6}$$

$$\Lambda = \Pi_{xu}, \tag{4.7}$$

$$\Pi = \Pi_{xs}. \tag{4.8}$$

The matrix L is the random function $AM_{uu}^{-1} + \Pi_{xs}M_{su}M_{uu}^{-1} + \Lambda$ of A, P is the random function $BM_{ss}^{-1} + \Pi$ of B. Then

$$dL = (dA)C^{-1}, \tag{4.9}$$

$$dP = (dB)E^{-1}. \tag{4.10}$$

However

$$\sigma(a_{ik}, a_{jl}) = \alpha_{ijkl}, \tag{4.11}$$

$$\sigma(a_{ik}, b_{jl}) = \beta_{ijkl}, \tag{4.12}$$

$$\sigma(b_{ik}, b_{jl}) = \gamma_{ijkl}, \tag{4.13}$$

where α_{ijkl}, β_{ijkl}, γ_{ijkl} are the appropriate quantities κ_{abcd}, respectively. From these we may compute $\sigma(l_{ij}, l_{kl})$, $\sigma(l_{ij}, p_{kl})$, and $\sigma(p_{ij}, p_{kl})$, the elements of the asymptotic covariance matrix of the elements of L and P (which are asymptotically normally distributed by the above). These elements can be estimated consistently from the sample (the proof follows from Theorem 1).

4.2. *The asymptotic distribution of $\hat{\beta}$ and $\hat{\gamma}$ for constant normalization.* In this section we shall show that $\hat{\beta}$ and $\hat{\gamma}$ are asymptotically normally distributed (Theorem 3). In view of the above theorem on asymptotic distributions the intricate part of the proof is in obtaining the covariance matrix. First we shall demonstrate that the elements of νW are $o(1/\sqrt{T})$ in probability. Since Assumption I holds, it is sufficient to show that $s(P_{xs}M_{ss}P'_{xs})$ is $o(1/\sqrt{T})$ in probability. This means $d \mid P_{xs}M_{ss}P'_{xs} \mid = 0$, since each of the characteristic roots of $P_{xs}M_{ss}P'_{xs}$ except the smallest approaches a non-zero limit in probability.

For any matrix A, A_{ij} denotes the matrix obtained by deleting the i-th row and j-th column from A, and $A_{ik,jl}$ is the matrix obtained by deleting the i-th and k-th rows and the j-th and l-th columns. Let

$$A^{ij} = (-1)^{i+j} | A_{ij} |,$$
$$A^{ij,kl} = (-1)^{i+j+k+l+\epsilon} | A_{ik,jl} |,$$

where $\epsilon = 0$ if $(i - k)(j - l) > 0$, 1 otherwise when $i \neq k$, $j \neq l$. $A^{ij,kl} = 0$ if $i = k$ or $j = l$. In the rest of the paper we use the summation convention of tensor calculus for lower case indices; namely, that whenever a lower case letter appears as a superscript and a subscript in an expression, the corresponding terms are to be summed on that index.

In general

$$d\,|A| = A^{ij} da_{ij}. \tag{4.14}$$

We may consider $P_{xs}M_{ss}P'_{xs}$ as a random function of P_{xs}. Then

$$d(i,j\text{-th element of } P_{xs}M_{ss}P'_{xs}) = \pi_i^k e_{kl}\, dp_j^l + \pi_j^l e_{kl}\, dp_i^k. \tag{4.15}$$

However

$$(\Pi_{xs}E\Pi'_{xs})^{ij} = \rho^j\beta^i = \rho^i\beta^j, \tag{4.16}$$

where ρ^i is a factor of proportionality. Since $\beta\Pi_{xs} = 0$, we have $d\,|P_{xs}M_{ss}P'_{xs}| = 0$. Then it can be shown that $d(\hat{\Pi}_{xs}M_{ss}\hat{\Pi}'_{xs} - P_{xs}M_{ss}P'_{xs}) = 0$, where $\hat{\Pi}_{xs} = \left(I - \dfrac{W_{xx}\hat{\beta}'\hat{\beta}}{\hat{\beta}W_{xx}\hat{\beta}'}\right) P_{xs}$.

Let $\Theta = \Pi_{xs}E\Pi'_{xs}$ and $F = P_{xs}M_{ss}P'_{sx}$. We know that $\hat{\beta}_i = \hat{\rho}_J\hat{\Theta}^{iJ}$, where $\rho_J = 1/\rho^J$ (and the capital letter J indicates that there is not to be a sum on that index), and $\hat{\Theta} = \hat{\Pi}_{xs}M_{ss}\hat{\Pi}'_{xs}$. Hence

$$d\hat{\beta}^i = \rho_J d\hat{\Theta}^{iJ} + \Theta^{iJ} d\hat{\rho}_J. \tag{4.17}$$

However $\hat{\beta}^i\hat{\beta}^j\varphi_{ij} = 1$; therefore $\hat{\rho}_J = (\hat{\Theta}^{iJ}\hat{\Theta}^{kJ}\varphi_{ik})^{-\frac{1}{2}}$ From this it follows that

$$d\hat{\rho}_J = -(\hat{\rho}_J)^3\Theta^{iJ}\varphi_{ik}\, d\hat{\Theta}^{kJ} \tag{4.18}$$

From (4.14) we see $d\hat{\Theta}^{kJ} = \Theta^{kJ,\alpha\beta}d\hat{\theta}_{\alpha\beta}$. Therefore

$$d\hat{\beta}^i = \rho_J[\Theta^{iJ,\alpha\beta} - \beta^i\beta^l\Theta^{kJ,\alpha\beta}\varphi_{kl}]d\hat{\theta}_{\alpha\beta}. \tag{4.19}$$

Let us define $\psi_j = \beta^i\varphi_{ij}$. Let us multiply (4.19) by $\theta_{\gamma i}$ and ψ_i. We obtain

$$\begin{aligned} \theta_{\gamma i} d\hat{\beta}^i &= \rho_J\theta_{\gamma i}\Theta^{iJ,\alpha\beta}d\hat{\theta}_{\alpha\beta} \\ &= \rho_J\delta_\gamma{}^J\Theta^{\alpha\beta}d\hat{\theta}_{\alpha\beta} - \rho_J\Theta^{J\beta}d\hat{\theta}_{\gamma\beta} = -\beta^\alpha d\hat{\theta}_{\gamma\alpha}, \end{aligned} \tag{4.20}$$

$$\psi_i d\hat{\beta}^i = 0. \tag{4.21}$$

Let us simplify (4.20). We see that

$$\beta^\alpha d\hat{\theta}_{\gamma\alpha} = \beta^\alpha\pi_\gamma^k e_{kl} dp_\alpha^l. \tag{4.22}$$

Hence

$$\sigma(\beta^\alpha d\hat{\theta}_{\gamma\alpha}, \beta^\mu d\hat{\theta}_{\gamma\mu}) = \beta^\alpha \pi^k_\gamma c_{kl}\beta^\mu \pi^h_\gamma c_{hj} c^{lm} c^{jt}\gamma_{\alpha\mu m t} = \beta^\alpha\beta^\mu\pi^m_\gamma\pi^t_\nu\gamma_{\alpha\mu m t} = r_{1\gamma\nu}, \tag{4.23}$$

say. Let $\sigma(\hat{\beta}^i, \hat{\beta}^j) = q_1^{ij}$, and let $Q_1 = (q_1^{ij})$. Then from (4.20) and (4.23) we obtain

$$\Theta Q_1 \Theta = R_1, \tag{4.24}$$

and (4.21) is

$$\psi Q_1 = 0. \tag{4.25}$$

It may be shown (see [1], for example) that the solution is

$$Q_1 = (I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(R_1)_{kk}(\Theta_{kk})^{-1}(I - \psi'\beta)_{k\cdot}, \tag{4.26}$$

where $k(1 \leq k \leq H)$ is arbitrary except that $\beta^k \neq 0$, and $A_{\cdot k}$ denotes A with the k-th column deleted, etc. If the normalization is $\beta^i = 1$, $k = i$ is a convenient choice.

Since $\hat{\gamma} = -\hat{\beta}L$,

$$d\hat{\gamma}^m = -d\hat{\beta}^i\lambda_i^m - \beta^i dl_i^m. \tag{4.27}$$

Hence

$$\sigma(\hat{\beta}^j, \hat{\gamma}^m) = -\sigma(\hat{\beta}^j, \hat{\beta}^i)\lambda_i^m - \sigma(\hat{\beta}^j, l_i^m)\beta^i, \tag{4.28}$$

$$\sigma(\hat{\gamma}^m, \hat{\gamma}^n) = \sigma(\hat{\beta}^j, \hat{\beta}^i)\lambda_i^m\lambda_j^n + \sigma(\hat{\beta}^j, l_i^m)\beta^i\lambda_j^n + \sigma(\hat{\beta}^j, l_i^n)\beta^i\lambda_j^m + \sigma(l_i^m, l_j^n)\beta^i\beta^j. \tag{4.29}$$

We, therefore, see that we must compute $\sigma(\hat{\beta}^j, l_i^m)\beta^i$ and $\sigma(l_i^m, l_j^n)\beta^i\beta^j$. We find, from (4.20), (4.21), and (4.22) that

$$\theta_{\gamma j}\beta^i\sigma(\hat{\beta}^j, l_i^m) = -\beta^i\beta^j\pi^k_j c^{mp}\beta_{ijpk} = r^m_{2\gamma}, \tag{4.30}$$

say. Let $(\sigma(\hat{\beta}^j, l_i^m)\beta^i) = Q_2$, and let $R_2 = (r^m_{2\gamma})$. Then, from (4.30) and (4.21) we obtain

$$\Theta Q_2 = R_2, \tag{4.31}$$

$$\psi Q_2 = 0. \tag{4.32}$$

The solution is

$$Q_2 = (I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(R_2)_{k\cdot}. \tag{4.33}$$

We find, readily, that

$$\beta^i\beta^j\sigma(l_i^m, l_j^n) = \beta^i\beta^j c^{mp}c^{nq}\alpha_{ijpq} = q_3^{mn}, \tag{4.34}$$

say, where $(c^{mp}) = C^{-1}$. Let $Q_3 = (q_3^{mn})$. This concludes the proof of Theorem 3.

THEOREM 3. *If Assumptions* A, B, F, H, I, K, L, *and* M *are satisfied*, $\sqrt{T}(\hat{\beta} - \beta)$ *and* $\sqrt{T}(\hat{\gamma} - \gamma)$ *are asymptotically jointly normally distributed with means zero and covariance matrix*

$$\sigma(\hat{\beta}', \hat{\beta}) = Q_1, \tag{4.35}$$

(4.36) $$\sigma(\hat{\beta}', \hat{\gamma}) = -Q_2\bar{\Pi}_{xu} - Q_2,$$

(4.37) $$\sigma(\hat{\gamma}', \hat{\gamma}) = \bar{\Pi}'_{xu}Q_1\bar{\Pi}_{xu} + \bar{\Pi}'_{xu}Q_2 + Q'_2\bar{\Pi}_{xu} + Q_3,$$

where Q_1 *is given by* (4.26), Q_2 *by* (4.33), *and* Q_3 *by* (4.34).

If there is a kind of asymptotic independence of ζ_t and z_t, then the above expressions may be simplified. Corollary 1 results from Theorem 3 and the following assumption:

ASSUMPTION N. $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{E}(\zeta_t^2 z'_t z_t) = \sigma^2 R$, *where* R *is defined in Assumption* F.

COROLLARY 1. *If Assumptions* A, B, F, H, I, K, L, M, *and* N *are satisfied,* $\sqrt{T}(\hat{\beta} - \beta)$ *and* $\sqrt{T}(\hat{\gamma} - \gamma)$ *are asymptotically jointly normally distributed with means zero and covariance matrix*

(4.38) $$\sigma(\hat{\beta}', \hat{\beta}) = \sigma^2(I - \beta'\psi)_{.k}(\Theta_{kk})^{-1}(I - \psi'\beta)_{k.},$$

(4.39) $$\sigma(\hat{\beta}', \hat{\gamma}) = -\sigma^2(I - \beta'\psi)_{.k}(\Theta_{kk})^{-1}(\bar{\Pi}_{xu} + \psi'\gamma)_{k.},$$

(4.40) $$\sigma(\hat{\gamma}', \hat{\gamma}) = \sigma^2[(\bar{\Pi}'_{xu} + \gamma'\psi)_{.k}(\Theta_{kk})^{-1}(\bar{\Pi}_{xu} + \psi'\gamma)_{k.} + C^{-1}].$$

4.3. *Asymptotic distribution of the estimates of the parameters* β *and* γ *with normalization a function of* Ω_{xx}.

If we relax Assumption L that Φ_{xx} is constant, we obtain a more general result. Since the proof, however, is more involved, we shall not give it here; the reader is referred to [1]. In the derivation of the estimates Ω_{xx} was defined as $\mathcal{E}(\delta'_t\delta_t)$. In the asymptotic theory we do not assume that this is the same for each t. We use the following assumption:

ASSUMPTION O.
$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{E}(\delta_{ti}\delta_{tj}\delta_{tk}z_{tl}) = n_{ijkl} \text{ exists;}$$
$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{E}(\delta_{ti}\delta_{tj}) = \bar{\omega}_{ij} \text{ exists;}$$
$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{E}(\delta_{ti}\delta_{tj}\delta_{tk}\delta_{tl}) = \bar{\omega}_{ijkl} + \bar{\omega}_{ij}\bar{\omega}_{kl} \text{ exists.}$$

Let δ_{ijkl} be the quantities n_{ijkl} corresponding to the u's, ϵ_{ijkl}, the quantities corresponding to the c's. Define

(4.41) $$\chi^{ij} = \tfrac{1}{2}\beta^k\beta^l \frac{\partial\varphi_{kl}}{\partial\omega_{ij}},$$

(4.42) $$r_{4\gamma} = \beta^k\pi^l_\gamma\chi^{ij}\epsilon_{ijkl},$$

(4.43) $$q'_4 = (I - \beta'\psi)_{.k}(\Theta_{kk})^{-1}(r'_4)_{k.},$$

(4.44) $$q_5 = \chi^{ij}\chi^{kl}\bar{\omega}_{ijkl},$$

(4.45) $$q^k_6 = \chi^{ij}\beta^m\delta_{ijml}c^{kl}.$$

With the aid of the matrices Q_1, Q_2, and Q_3, the vectors q_4 and q_6, and the

scalar q_5, we may express the asymptotic covariance matrix of the estimates. We obtain

THEOREM 4. *If Assumptions* A, B, F, H, I, K, M, *and* O *are satisfied, and* Φ_{zz} *is a function of* Ω_{zz}, $\sqrt{T}(\hat{\beta} - \beta)$ *and* $\sqrt{T}(\hat{\gamma} - \gamma)$ *are asymptotically jointly normally distributed with means zero and covariance matrix*

$$\sigma(\hat{\beta}', \hat{\beta}) = Q_1 + q_4'\beta + \beta' q_4 + q_5\beta'\beta, \tag{4.46}$$

$$\sigma(\hat{\beta}', \hat{\gamma}) = -Q_1\bar{\Pi}_{zu} + q_4'\gamma - \beta' q_4\bar{\Pi}_{zu} + q_5\beta'\gamma - Q_2 - \beta' q_6, \tag{4.47}$$

$$\begin{aligned}\sigma(\hat{\gamma}', \hat{\gamma}) = {} & \bar{\Pi}_{zu}' Q_1 \bar{\Pi}_{zu} - \bar{\Pi}_{zu}' q_4'\gamma - \gamma' q_4 \bar{\Pi}_{zu} + q_5\gamma'\gamma \\ & + \bar{\Pi}_{zu}' Q_2 + Q_2'\bar{\Pi}_{zu} - \gamma' q_6 - q_6'\gamma + Q_3,\end{aligned} \tag{4.48}$$

where Q_1, Q_2, Q_3, q_4, q_5, *and* q_6 *are given by* (4.26), (4.33), (4.34), (4.43), (4.44), *and* (4.45) *respectively.*

COROLLARY 2. *If Assumptions* A, B, D, F, H *and* K *are satisfied, and* $\Phi_{zz} = \Omega_{zz}$, $\sqrt{T}(\hat{\beta} - \beta)$ *and* $\sqrt{T}(\hat{\gamma} - \gamma)$ *are asymptotically jointly normally distributed with means zero and covariance matrix*

$$\sigma(\hat{\beta}', \hat{\beta}) = (I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(I - \psi'\beta)_{k\cdot} + \tfrac{1}{2}\beta'\beta, \tag{4.49}$$

$$\sigma(\hat{\beta}', \hat{\gamma}) = -(I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(\bar{\Pi}_{zu} + \psi'\gamma)_{k\cdot} + \tfrac{1}{2}\beta'\gamma, \tag{4.50}$$

$$\sigma(\hat{\gamma}', \hat{\gamma}) = (\bar{\Pi}_{zu}' + \gamma'\psi)_{\cdot k}(\Theta_{kk})^{-1}(\bar{\Pi}_{zu} + \psi'\gamma)_{k\cdot} + C^{-1} + \tfrac{1}{2}\gamma'\gamma. \tag{4.51}$$

5. Asymptotic distribution of the likelihood ratio criterion and the small sample criterion for testing a certain hypothesis. The likelihood ratio criterion for testing the hypothesis that the number of coordinates of z_t with zero coefficients in the selected structural equation is as great as it is assumed to be is $(1 + \nu)^{-\frac{1}{2}T}$ [2, Theorem 2], where ν is the smallest root of

$$| P_{zz}M_{zz}P_{zz}' - \nu W_{zz} | = 0. \tag{5.1}$$

Then

$$T\nu = T\frac{\hat{\beta}P_{zz}M_{zz}P_{zz}'\hat{\beta}'}{\hat{\beta}W_{zz}\hat{\beta}'} = (\sqrt{T}\hat{\beta}P_{zz})\frac{M_{zz}}{\hat{\beta}W_{zz}\hat{\beta}'}(\sqrt{T}\hat{\beta}P_{zz})'. \tag{5.2}$$

From Theorem 5 of [4] it follows that the asymptotic distribution of $T\nu$ is the same as that of the quadratic form $x\dfrac{E}{\sigma^2}x'$, where x has the limiting distribution of $\sqrt{T}\hat{\beta}P_{zz}$, use being made of $\operatorname*{plim}_{T\to\infty} \hat{\beta}W_{zz}\hat{\beta}' = \sigma^2$. We have

$$dx^i = \beta^j dp_j^i + d\hat{\beta}^j\pi_j^i. \tag{5.3}$$

Let $\Upsilon = (I - \beta'\psi)_{\cdot k}(\Theta_{kk})^{-1}(I - \psi'\beta)_{k\cdot}$. Then

$$d\hat{\beta}^j = -\upsilon^{jk}\beta^l\pi_k^m e_{mn}dp_l^n. \tag{5.4}$$

Substituting in (5.3), we obtain

$$dx^i = \beta^j dp_j^i - \upsilon^{lk}\beta^j\pi_k^m e_{mn}dp_j^n\pi_l^i. \tag{5.5}$$

Then

$$\sigma(x^i, x^g) = \sigma^2(e^{ig} - \pi_k^i \pi_q^g v^{kq}) = \sigma^2 \xi^{ig} \tag{5.6}$$

say, and $(\xi^{ig}) = \Xi$

Let F be chosen so $E = FF'$ and $F'\Xi F = \Psi$ is diagonal. Since $E\Xi E\Xi E = E\Xi E$, the diagonal elements of Ψ are 1 and 0. The number of elements that are 1 is the rank of $E\Xi E$, namely, $D - H + 1$, where D is the number of coordinates of v_t (the number of coordinates whose coefficients in the selected equation are assumed to be zero). Let $z = \frac{1}{\sigma} xF$. Then the asymptotic distribution of $T\nu$ is the distribution of zz' where z is normally distributed with mean zero and covariance matrix Ψ. It is the χ^2-distribution with $D - H + 1$ degrees of freedom We observe that $T \log (1 + \nu)$ and $TD\lambda$ are asymptotically equal to $T\nu$, where λ is the criterion based on small sample theory [2, Theorem 4]. Finally, we note that ν is independent of the normalization of β.

THEOREM 5. *If Assumptions* A, B, F, H, I, K, M, *and* N *are satisfied, -2 times the logarithm of the likelihood ratio criterion, $-T/2 \log (1 + \nu)$, the asymptotically equivalent $T\nu$ and TD times the small sample criterion, λ, for testing the hypothesis that the number of coordinates with zero coefficients is D are asymptotically distributed as χ^2 with $D - H + 1$ degrees of freedom.*

This theorem indicates how conservative the small sample test is asymptotically, for that test asymptotically is equivalent to using $T\nu$ as having an asymptotic χ^2-distribution with D degrees of freedom.

6. Asymptotic behavior of confidence regions based on small sample theory. In [2] we deduced confidence regions for β and for β and γ when Assumption E holds. If the normalization of β is

$$\beta \Phi_{xx} \beta' = 1, \tag{6.1}$$

where Φ_{xx} is a given matrix, then a confidence region (a) for β of confidence ϵ consists of all β^* satisfying (6.1) and

$$\frac{\beta^* M_{xz} M_{zz}^{-1} M_{zx} \beta^{*\prime}}{\beta^* W_{xx} \beta^{*\prime}} \leq \frac{D}{T - K} F_{D, T-K}(\epsilon), \tag{6.2}$$

where $F_{D,T-K}(\epsilon)$ is chosen so the probability of (6.2) for $\beta^* = \beta$ is ϵ and K is the number of coordinates of z_t and D is the number of coordinates of v_t A region (b) for β and γ simultaneously consists of β^* and γ^* satisfying (6.1) and

$$\frac{\beta^* M_{xu} M_{uu}^{-1} M_{ux} \beta^{*\prime} + \beta^* M_{xu} \gamma^{*\prime} + \gamma^* M_{ux} \beta^{*\prime} + \gamma^* M_{uu} \gamma^{*\prime} + \beta^* M_{xz} M_{zz}^{-1} M_{zx} \beta^{*\prime}}{\beta^* W_{xx} \beta^{*\prime}} \leq \frac{K}{T - K} F_{K, T-K}(\epsilon). \tag{6.3}$$

We shall now show that even if Assumption E does not hold the regions have asymptotically confidence coefficients ϵ and they are consistent under general conditions.

Let $c = \beta M_{xu} M_{uu}^{-1} + \gamma$, $e = \beta M_{xs} M_{ss}^{-1}$. We observe from Section 4 that if Assumptions A, B, F, H, K, L, M and N are satisfied, the vectors $\sqrt{T}c$ and $\sqrt{T}e$ have asymptotic independent distributions $N(0, \sigma^2 C^{-1})$ and $N(0, \sigma^2 E^{-1})$, respectively. Then $TcM_{uu}c'/\sigma^2$ and $TeM_{ss}e'/\sigma^2$ will have asymptotic independent χ^2-distributions with $F(= K - D)$ and D degrees of freedom, respectively. Also $\beta W_{xx}\beta'$ approaches σ^2 stochastically. By Theorems 5 and 6 of [4], the left-hand sides of (6.2) and (6.3) have asymptotic F-distributions with D and $T - K$ degrees of freedom and K and $T - K$ degrees of freedom, respectively.

We shall prove that (a) is consistent for β; the proof is similar for (b) as a region for β and γ. If we replace β by b in the definition of e, $eM_{ss}e' = bM_{xs}M_{ss}^{-1}M_{sx}b'$. For $b \neq \beta$ we must show that the probability that b will fall in the confidence region for β approaches zero. The above form approaches $b\Pi_{xs}E\Pi_{xs}'b'$ in probability. If $b \neq \beta$ and satisfies (6.1) then $b\Pi_{xs} \neq 0$ and $eM_{ss}e'$ has a non-zero limit in probability since E is positive definite. Thus b is not in the limiting confidence region.

THEOREM 6. *If Assumptions* A, B, F, H, I, K, M, *and* N *are satisfied, the confidence regions of Theorem* 3 *of* [2] (*including* (a) *and* (b) *above*) *are consistent, and the regions* (a) *and* (b) *have asymptotically the confidence levels* ϵ.

REFERENCES

[1] T. W. ANDERSON AND HERMAN RUBIN, "Estimation of the parameters of a single stochastic difference equation in a complete system," Cowles Commission for Research in Economics, 1947, dittoed.

[2] T. W. ANDERSON AND HERMAN RUBIN, "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Math. Stat*, Vol. 20 (1949), pp. 46–63.

[3] H. RUBIN, "Consistency and asymptotic normality in stable linear stochastic difference systems," to be published.

[4] H. RUBIN, "Topological properties of measures on topological spaces," *Duke Math Journ.*, to be published.

SOME NONPARAMETRIC TESTS OF WHETHER THE LARGEST OBSERVATIONS OF A SET ARE TOO LARGE OR TOO SMALL

By John E. Walsh

The Rand Corporation

1. Summary. Let us consider a large number n of observations which are statistically independent and drawn from continuous symmetrical populations. This paper presents some nonparametric tests of whether the r largest observations of the set are too large to be consistent with the hypothesis that these populations have a common median value. Tests of whether the r largest observations are too small to be consistent with this hypothesis are also considered. Here r is a given integer which is independent of n.

Subject to some weak restrictions, it is shown that the significance level of a test of the type presented tends to a value α as n increases. For no admissible value of n, however, does the significance level of this test exceed 2α. If whether the largest observations are too large is considered, tests with values of α suitable for significance levels can be obtained for $r \geq 4$. Values of α suitable for significance levels can be obtained for any value of r if whether the largest observations are too small is investigated (n large).

Properties of the power functions of these tests are considered for the special case in which the r largest observations are from populations with common median θ, the remaining observations are from populations with common median ϕ, and each population has the property that the distribution of the quantity

$$\text{(sample value)} - \text{(population median)}$$

is independent of the value of the population median. For tests of $\theta > \phi$, the power function tends to zero as $\theta - \phi \to -\infty$ and to unity as $\theta - \phi \to \infty$. For tests of $\phi > \theta$, the power function tends to unity as $\theta - \phi \to -\infty$ and to zero as $\theta - \phi \to \infty$.

Analogous tests of whether the smallest observations of a set are too small or too large can be obtained from the tests of the largest observations by symmetry considerations.

If there is strong reason to believe that the set of observations is a random sample from a continuous population, the tests presented in this paper can be used to decide whether the population is symmetrical. Tests of this nature are sensitive to symmetry in the tails of the population but not to symmetry in the central part.

2. Introduction and statement of tests. The tests derived in this paper are applicable to situations of the following two types:

(a). It is known that the observations are independent and from continuous

symmetrical populations (i.e., each population has a continuous cdf $F(x)$ such that $F(x - \phi) = 1 - F(\phi - x)$, where ϕ is the population median). It is desired to test whether the largest few observations are too large (or too small) to be consistent with the assumption that the populations have a common median value (if the 50% point of a continuous symmetrical population is not unique, the median of this population is *defined* to be the midpoint of the interval of 50% points).

(b). It is known that the observations are independent and from continuous populations with a common median value (e.g., the observations may be a sample from a continuous population). It is desired to test whether these populations are symmetrical (with emphasis on the tails of the population).

With respect to (a), perhaps the most common practical application is that where the observations are assumed to be a sample from a continuous symmetrical population of some special type (e.g., normal) but the values of the largest few observations make this assumption questionable. The nonparametric tests presented for (a) are easily applied and a significant result for a nonparametric test automatically implies that the observations are not a sample from the specified type of population. Furthermore, if a parametric test of this situation (i.e., a test based on the assumption of a sample from this special type of population) is significant, the nonparametric tests are useful in determining whether it is possible that the observations might be a sample from a continuous symmetrical population of some other type.

With respect to (b), perhaps the most common application is that where the set of observations can be considered to be a sample from a continuous population and it is desired to test whether this population is symmetrical in the tails.

Now let us consider the forms of the tests. Let $x(1), \cdots, x(n)$ represent the values of the n observations arranged in increasing order of magnitude. Then $x(n + 1 - r), x(n + 2 - r), \cdots, x(n)$ are the r largest observations of the set. For situations of type (a), the tests of whether the r largest observations are too large are of the form

Test 1. *Accept that the r largest observations are too large to be consistent with the hypothesis that the populations have a common median if*

$$\min [x(n + 1 - i_k) + x(j_k); 1 \leq k \leq s \leq r] > 2x(W_\alpha),$$

where the i's, j's and n are integers such that

$$i_s = r, \qquad i_u < i_{u+1}, \qquad j_v < j_{v+1}, \qquad j_s < W_\alpha < n + 1 - r,$$

α is defined by

$$\alpha = Pr\{\min [x(n + 1 - i_k) + x(j_k)] > 2\phi \mid \phi = \text{common median}\},$$

and $W_\alpha = W_\alpha(n)$ is the smallest integer satisfying the relation

$$Pr[x(W_\alpha) < \phi \mid \phi = \text{common median}] \leq \alpha. \tag{1}$$

In testing the hypothesis of Test 1, the principle followed is to choose $x(n + 1 - r)$ and some subset of $x(n + 2 - r), \cdots, x(n)$ for use in the test. The integer s represents the total number of order statistics selected from $x(n + 1 - r), \cdots, x(n)$.

The value of $\alpha = \alpha(i_1, \cdots, i_s; j_1, \cdots, j_s)$ is independent of n and is given by equation (4) in Section 3. Table 1 contains some values of the i's, j's and s which yield values of α suitable for significance levels. For Test 1, values of α suitable for significance levels can be obtained for $r \geq 4$.

TABLE 1

Some values of α for $s \leq 5$

α	s	i_1	i_2	i_3	i_4	i_5	j_1	j_2	j_3	j_4	j_5
.0625	1	4					1				
.0312	1	5					1				
.0156	1	6					1				
.0078	1	7					1				
.0039	1	8					1				
.0352	1	7					2				
.0195	1	8					2				
.0107	1	9					2				
.0469	2	4	5				1	2			
.0234	2	5	6				1	2			
.0117	2	6	7				1	2			
.0059	2	7	8				1	2			
.0391	3	4	5	6			1	2	3		
.0195	3	5	6	7			1	2	3		
.0098	3	6	7	8			1	2	3		
.0459	4	4	5	6	7		1	2	3	4	
.0229	4	5	6	7	8		1	2	3	4	
.0115	4	6	7	8	9		1	2	3	4	
.0308	5	4	5	6	7	8	1	2	3	4	5
.0154	5	5	6	7	8	9	1	2	3	4	5
.0077	5	6	7	8	9	10	1	2	3	4	5

If the n independent observations satisfy the additional conditions

(A)
(i). Asymptotically $(n \to \infty)$, $x(W_\alpha)$ is statistically independent of min $[x(n + 1 - i_k) + x(j_k); 1 \leq k \leq s]$.
(ii). The standard deviations of $x(W_\alpha)$ and min $[x(n + 1 - i_k) + x(j_k); 1 \leq k \leq s]$ exist for all $n \geq i_s + j_s - 1$ and the limiting ratio $(n \to \infty)$ of these standard deviations is either zero or infinite.
(iii). Let the notation $\sigma(z)$ denote the standard deviation of z Then, if the populations have a common median ϕ, asymptotically the *cdf*'s of

$[x(W_\alpha) - \phi]/\sigma[x(W_\alpha)]$ and $\{\min [x(n + 1 - i_k) + x(j_k)] - 2\phi\}/\sigma\{\min [x(n + 1 - i_k) + x(j_k)]\}$ are continuous at the point zero.

then the significance level of Test 1 approaches the value α as n tends to infinity.

Although conditions (A) may appear to be complicated, they are not very restrictive. These conditions are satisfied if the n observations are a sample from a continuous population of the type usually encountered in practical situations (i.e., approximated in practical situations). Perhaps the most well known type of continuous symmetrical population for which a sample does not satisfy conditions (A) is that with a triangular probability density function. Part (ii) of conditions (A) is not satisfied for a sample from a population of this type.

For large n, relation (1) with the equality sign is approximately satisfied if $W_\alpha = \frac{1}{2}n + \frac{1}{2}K_\alpha\sqrt{n}$, (i.e., the largest integer contained in $\frac{1}{2}n + \frac{1}{2}K_\alpha\sqrt{n}$). Here K_α is the standardized normal deviate exceeded with probability α. This value for W_α was obtained from the normal approximation to the binomial theorem and furnishes a reasonably accurate solution of (1) with the equality sign for $n > 10$, (see [1]).

As an example of a test of type 1, let $r = 5$, $s = 2$, $j_1 = 1$, $j_2 = 2$, $i_1 = 4$, $i_2 = 5$. Then $\alpha \doteq .0547$ and the test is (approximately)

TEST 2. *Accept the specified alternative of Test* 1 *if*

$$\min [x(n - 3) + x(1), x(n - 4) + x(2)] > 2x(\tfrac{1}{2}n + \tfrac{1}{2}K_{.0547}\sqrt{n}).$$

That this is a test of whether the 5 largest observations are too large is intuitively evident from the fact that a significant result will be obtained only if both

$$\begin{aligned} x(n - 3) &> 2x(\tfrac{1}{2}n + \tfrac{1}{2}K_{.0547}\sqrt{n}) - x(1), \\ x(n - 4) &> 2x(\tfrac{1}{2}n + \tfrac{1}{2}K_{.0547}\sqrt{n}) - x(2). \end{aligned} \tag{2}$$

If the smallest two of the five largest observations are too large, it seems reasonable to suppose that all of the five are too large A similar interpretation exists for all tests of the type of Test 1.

The type (a) tests of whether the largest observations are too small are of the form

TEST 3. *Accept that the r largest observations are too small to be consistent with the hypothesis that the populations have a common median value if*

$$\max [x(n + 1 - j_k) + x(i_k); 1 \le k \le s \le r] < 2x(n + 1 - W_\alpha),$$

where $j_s = r$, $j_v < j_{v+1}$, $i_u < i_{u+1}$, $i_s < n + 1 - W_\alpha < n + 1 - r$, and both α and W_α are defined in Test 1.

From the results for Test 1 and symmetry considerations, the significance level of test 3 tends to α as $n \to \infty$ if conditions (A) are satisfied; it does not exceed 2α for any admissible value of n. For Test 3, values of α suitable for significance levels can be obtained for all values of r (n sufficiently large).

As indicated by (2), the tests of whether the largest observations are too large

can also be interpreted as tests of whether the smallest observations are too large. Similarly the tests of whether the largest observations are too small can also be interpreted as tests of whether the smallest observations are too small.

The above discussion presents intuitive reasons for believing that Tests 1 and 3 are suitable for the situations to which they are applied. To obtain a semi-quantitative measure of the suitability of these tests, this paper investigates the special case in which the r largest observations are from continuous symmetrical populations with common median θ, the remaining observations are from continuous symmetrical populations with common median ϕ, and each population has the property that the distribution of $x - \psi$ is independent of ψ, where x is an observation from the population and ψ is the median of the population. The power function of a test of type 1 or 3 is defined to be the probability that the test is significant given the value of $\theta - \phi$. It is found that the power functions of these tests have several desirable properties: For Test 1, the power function tends to zero as $\theta - \phi \to -\infty$, is a monotonically increasing function of $\theta - \phi$ for $\theta - \phi < 0$, and tends to unity as $\theta - \phi \to \infty$. For Test 3, the power function tends to zero as $\theta - \phi \to \infty$, is monotonically decreasing for $\theta - \phi < 0$, and tends to unity as $\theta - \phi \to -\infty$.

For testing whether the populations are symmetrical in the tails given that they are continuous and have a common median, i.e , situation (b), a combination of 1 and 3 is used. The resulting test is

TEST 4. *Accept that the populations are not symmetrical in the tails if either*

$$\min [x(n + 1 - i_k) + x(j_k); 1 \leq k \leq s] > 2x(W_\alpha)$$

or

$$\max [x(n + 1 - j_k) + x(i_k); 1 \leq k \leq s] < 2x(n + 1 - W_\alpha),$$

where $\alpha < \frac{1}{2}$, $i_u < i_{u+1}$, $j_v < j_{v+1}$, $j_w \leq i_w$, $j_s < W_\alpha < n + 1 - i_s$, *and both* α *and* W_α *are defined in Test* 1.

Since both inequalities in Test 4 can not be satisfied simultaneously, the significance level of Test 4 tends to 2α as $n \to \infty$ if conditions (A) are satisfied; it never exceeds 4α for any admissible value of n.

The asymptotic distribution ($n \to \infty$) of $x(W_\alpha)$ is usually not very sensitive to symmetry of the populations. For example, if the n observations are a sample from a population with a probability density function $f(x)$ such that ($f(\phi) \neq 0$, (ϕ = population 50% point), and $f'(x)$ exists and is continuous in a neighborhood of $x = \phi$, it can be shown that the only property of $f(x)$ which influences the asymptotic distribution of $x(W_\alpha)$ is the value of $f(\phi)$. Thus, since a type 1 test investigates both whether the largest observations are too large and whether the smallest observations are too large (to be consistent with the assumption of symmetry), while a type 3 test investigates both whether the largest observations are too small and whether the smallest observations are too small, Test 4 should be suitable for testing whether a population has symmetrical tails

3. Theorems and derivations. The fundamental fact used in this paper is that, if the observations are from continuous symmetrical populations with common median ϕ, the value of

$$\alpha = Pr\{\min [x(n + 1 - i_k) + x(j_k); 1 \le k \le s] > 2\phi\}$$
$$= Pr\{\max [x(n + 1 - j_k) + x(i_k); 1 \le k \le s] < 2\phi\}$$

is independent of n for the values of n permitted in the tests. This result is a special case of the following theorem

THEOREM 1. *Consider a set of n independent observations from continuous symmetrical populations with common median ϕ. Let $i_1 < \cdots < i_s$ and $j_1 < \cdots < j_s$ be fixed sets of integers whose values are independent of n. Then the value of*

$$Pr\{\beta\text{th largest of } [x(n + 1 - j_k) + x(i_k); 1 \le k \le s] < 2\phi\}$$

is the same for all values of n which are $\ge i_s + j_s - 1$. In particular

$$\alpha = 2^{-w}\Big\{1 + m(1) + \sum_{h_1=1}^{m(2)} [m(1) - h_1] + \sum_{h_2=1}^{m(3)} \sum_{h_1=1}^{m(2)-h_2} [m(1) - h_1 - h_2] + \cdots + \sum_{h_{u-1}=1}^{m(u)} \sum_{h_{u-2}=1}^{m(u-1)-h_{u-1}} \cdots \sum_{h_1=1}^{m(2)-h_2-\cdots-h_{u-1}} [m(1) - h_1 - \cdots - h_{u-1}]\Big\}, \tag{3}$$

where

$$w = i_s + j_s - 1, \quad u = j_s - 1, \quad m(j_t + v_t - 1) = i_s + j_s - i_t - j_t - v_t + 1,$$
$$t = 0, 1, \cdots, s - 1, \qquad 1 \le v_t \le j_{t+1} - j_t, \qquad i_0 = j_0 - 1 = 0.$$

PROOF. It is sufficient to prove the theorem for the expression

$$Pr\{\max [x(n + 1 - j_k) + x(i_k); 1 \le k \le s] < 2\phi\},$$

since any probability expression of the form $Pr\{\beta\text{th largest of } [\ ] < 2\phi\}$ can be expressed as a specified constant plus a sum of probabilities of the form $Pr\{\max [\ ] < 2\phi\}$ multiplied by specified constants, where in each case the terms in the $[\ ]$ are a subset of the s terms: $x(n + 1 - j_k) + x(i_k)$, $(1 \le k \le s)$.

Let the integer n have the value n_0. Then it can be verified that

$$\begin{aligned} &Pr\{\max [x(n_0 + 1 - j_k) + x(i_k); 1 \le k \le s] < 2\phi\} \\ &\quad = Pr[\max \{2x(n_0 - j_s), x[n_0 + 1 - W] + x[n_0 + 1 - W - m(W)]; \\ &\qquad\qquad 1 \le W \le j_s\} < 2\phi], \end{aligned} \tag{4}$$

where

$$m(j_t + v_t - 1) = n_0 + 2 - i_t - j_t - v_t, \qquad m(j_s) = n_0 - i_s - j_s \ge 1,$$
$$t = 0, 1, \cdots, s - 1, \qquad 1 \le v_t \le j_{t+1} - j_t, \qquad i_0 = j_0 - 1 = 0,$$

by the use of Theorem 4 of [2]. By the proof of Theorem 5 of [2], the value of the second term in (4) equals

$$Pr[\max \{2x(n_0 - j_s), x[n_0 + 2 - W] + x[n_0 + 1 - W - m(W)],$$
$$1 \le W \le j_s + 1\} < 2\phi]$$

if $m(j_s + 1) = 1$ and the expression is based on $n_0 + 1$ rather than n_0 observations (the values of the m's are the same as in (4)). The value of this expression, however, can be shown to equal the value of

$$Pr[\max \{2x(n_0 + 1 - j_s), x[n_0 + 2 - W] + x[n_0 + 2 - W - m(W)];$$
$$1 \le W \le j_s\} < 2\phi],$$

which by (4) equals the value of

$$Pr\{\max [x(n_0 + 2 - j_k) + x(i_k); 1 \le k \le s] < 2\phi\}$$

if $n = n_0 + 1$ for this expression. Thus, by induction, the value of

$$Pr\{\max [x(n + 1 - j_k) + x(i_k); 1 \le k \le s] < 2\phi\}$$

is the same for all sample sizes $n \ge i_s + j_s$. An analysis similar to that used in the proof of Theorem 5 of [2] shows that this also holds for $n = i_s + j_s - 1$. Equation (3) was obtained by taking $n = w = i_s + j_s - 1$, the m's as given by (4) with this value of n, and substituting into Theorem 4 of [2]

Another basic result is that, if the observations are from continuous symmetrical populations with common median ϕ, the value of

$$Pr\{\min [x(n + 1 - i_k) + x(j_k), 1 \le k \le s] > 2x(W_\alpha)\}$$
$$= Pr\{\max [x(n + 1 - j_k) + x(i_k); 1 \le k \le s] < 2x(n + 1 - W_\alpha)\}$$

is always less than or equal to 2α. This is a particular application of the theorem

THEOREM 2. *Consider n independent observations from continuous symmetrical populations with common median ϕ. Then, for any integer W,*

$$Pr\{\max [x(n + 1 - j_k) + x(i_k); 1 \le k \le s] < 2x(W)\}$$
$$\le Pr\{\max [x(n + 1 - j_k) + x(i_k)] < 2\phi\} + Pr\{x(W) > \phi\}$$
$$- Pr\{\max [x(n + 1 - j_k) + x(i_k)] < 2\phi, x(W) > \phi\}.$$

PROOF.

$$Pr\{\max [\ ] < 2x(W)\} = Pr\{\max [\ ] < 2\phi, x(W) > \phi\}$$
$$+ Pr\{\max [\ ] < 2\phi, x(W) < \phi, \max [\ ] < 2x(W)\}$$
$$+ Pr\{\max [\ ] > 2\phi, x(W) > \phi, \max [\ ] < 2x(W)\}$$
$$\le Pr\{\max [\ ] < 2\phi, x(W) > \phi\} + Pr\{\max [\ ] < 2\phi, x(W) < \phi\}$$
$$+ Pr\{\max [\ ] > 2\phi, x(W) > \phi\}$$
$$= Pr\{\max [\ ] < 2\phi\} + Pr\{x(W) > \phi\} - Pr\{\max [\ ] < 2\phi, x(W) > \phi\}.$$

If the n independent observations satisfy conditions (A) in addition to being from continuous symmetrical populations with a common median value, the significance level of Tests I and 3 tends to α as $n \to \infty$. This follows from symmetry considerations and

THEOREM 3. *Consider n independent observations which satisfy conditions* (A) *and are from continuous symmetrical populations with a common median value. Then*

$$\lim_{n\to\infty} Pr\{\min [x(n + 1 - i_k) + x(j_k); 1 \leq k \leq s] > 2x(W_\alpha)\} = \alpha.$$

PROOF. Let

$$Y = \min [x(n + 1 - i_k) + x(j_k); 1 \leq k \leq s]$$

and consider the case where

$$\lim_{n\to\infty} \sigma[x(W_\alpha)]/\sigma(Y) = 0.$$

Since the populations are continuous, $\sigma(Y) > 0$ and

$$\begin{aligned} Pr[Y > 2x(W_\alpha)] &= Pr[Y - 2\phi > 2x(W_\alpha) - 2\phi] \\ &= Pr\{[Y - 2\phi]/\sigma(Y) > 2[x(W_\alpha) - \phi]/\sigma(Y)\}. \end{aligned}$$

Let

$$Z = 2[x(W_\alpha) - \phi]/\sigma(Y).$$

Then, from (i) of conditions (A),

$$Pr[Y > 2x(W_\alpha)] = \int_{-\infty}^{\infty} Pr\{[Y - 2\phi]/\sigma(Y) > a\}\, dF_z(a) + \beta(n),$$

where F_z is the *cdf* of Z and $\lim_{n\to\infty} \beta(n) = 0$.

Let b be any positive number. From $\lim_{n\to\infty} \sigma(Z) = 0$, (ii) of conditions (A), and the definition of $x(W_\alpha)$, the mean of Z exists for all values of n and tends to zero as $n \to \infty$. Then, by Tchebycheff's Inequality, it can be shown that

$$\int_{-b}^{b} dF_z(a) = 1 - \gamma(n),$$

where $\lim_{n\to\infty} \gamma(n) = 0$.

From (iii) of conditions (A)

$$\lim_{n\to\infty} Pr\{[Y - 2\phi]/\sigma(Y) > -b\} = \lim_{n\to\infty} Pr\{[Y - 2\phi]/\sigma(Y) > b\} + \delta(b),$$

where $\lim_{b\to 0} \delta(b) = 0$.

Using the above relations, letting $n \to \infty$ first and then $b \to 0$, it follows from Theorem 1 that

$$\lim_{n\to\infty} Pr[Y > 2x(W_\alpha)] = Pr\{[Y - 2\phi]/\sigma(Y) > 0\} = \alpha.$$

A similar type proof shows that this limiting relation also holds when

$$\lim_{n\to\infty} \sigma[x(W_\alpha)]/\sigma(Y) = \infty.$$

Finally consider properties of the power functions of Tests 1 and 3 for the special situation outlined in sections 1 and 2. The properties stated in the preceding two sections follow from

THEOREM 4. *Let $x(n+1-r), \cdots, x(n)$ be from continuous symmetrical populations with common median θ, the remaining order statistics from continuous symmetrical populations with common median ϕ, and each population have the property that the distribution of $x - \psi$ is independent of ψ, where x is an observation from the population and ψ is the median of the population. Also let*

$$P_1(\Phi) = Pr\{\min [x(n+1-i_k) + x(j_k); 1 \le k \le s \le r] > 2x(W_\alpha) \mid \theta - \phi = \Phi\},$$

where the conditions for Test 1 are satisfied, and

$$P_3(\Phi) = Pr\{\max [x(n+1-j_k) + x(i_k); 1 \le k \le s \le r] < 2x(n+1-W_\alpha) \mid \theta - \phi = \Phi\},$$

where the conditions for Test 3 are satisfied. Then

$$\lim_{\Phi\to-\infty} P_1(\Phi) = 0, \qquad \lim_{\Phi\to\infty} P_1(\Phi) = 1,$$

$$\lim_{\Phi\to-\infty} P_3(\Phi) = 1, \qquad \lim_{\Phi\to\infty} P_3(\Phi) = 0,$$

$P_1(\Phi)$ is a monotonically increasing function of Φ for $\Phi < 0$, and $P_3(\Phi)$ is a monotonically decreasing function of Φ for $\Phi < 0$

PROOF. It is sufficient to prove this theorem for the power function of Test 3. The results for $P_1(\Phi)$ can be obtained from symmetry considerations and obvious modifications of the proof for $P_3(\Phi)$.

First consider $P_3(\Phi)$ for the case where $\Phi \le 0$. Let a new set of observations be formed from the given set by subtracting the median value of the corresponding population from each observation. Let $y(1), \cdots, y(n)$ be the values of the set of modified observations arranged in increasing order of magnitude. Since $\Phi \le 0$, $\theta \le \phi$ and

$$y(t) = \begin{cases} x(t) - \phi, & 1 \le t \le n - r, \\ x(t) - \theta, & n - r + 1 \le t \le n. \end{cases}$$

Thus

$$P_3(\Phi) = Pr\{\max [y(n+1-j_k) + y(i_k); 1 \le k \le s \le r] - 2y(n+1-W_\alpha) < -\Phi\},$$

whence it follows that $P_3(\Phi)$ is a monotonically decreasing function of Φ for $\Phi \leq 0$ and that $\lim_{\Phi \to -\infty} P_3(\Phi) = 1$.

Now consider the case where $\Phi > 0$. Again form the set of modified observations and let $y(1), \cdots, y(n)$ be the values of these observations arranged in increasing order of magnitude. Then it is easily seen that

$$P_3(\Phi) \leq Pr[y(1) - y(n) < -\tfrac{1}{2}\Phi]$$

so that $\lim_{\Phi \to \infty} P_3(\Phi) = 0$.

REFERENCES

[1] Paul G Hoel, *Introduction to Mathematical Statistics*, John Wiley and Sons, 1947, p. 45.

[2] John E. Walsh, "Some significance tests for the median which are valid under very general conditions," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 64–81.

ON A MEASURE OF DEPENDENCE BETWEEN TWO RANDOM VARIABLES

By Nils Blomqvist

University of Stockholm and Boston University

1. Summary. The properties of a measure of dependence q' between two random variables are studied. It is shown (Sections 3–5) that q' under fairly general conditions has an asymptotically normal distribution and provides approximate confidence limits for the population analogue of q' A test of independence based on q' is non-parametric (Section 6), and its asymptotic efficiency in the normal case is about 41% (Section 7). The q'-distribution in the case of independence is tabulated for sample sizes up to 50.

2. Introduction and definitions. In drawing conclusions from statistical data it frequently happens that it is unnecessary to utilize all the information given by the data. In such cases it seems desirable to use methods which are

1) valid under rather weak assumptions regarding the distribution of the population and

2) easy to deal with in practice.

Naturally such methods should always be used, but their applicability is, in most cases, limited by their small efficiency.

Concerning methods of measuring correlation and testing independence some so-called rank correlation coefficients have been defined [2, 3, 4, 6] which have the first property. In large samples these are, however, rather tiresome to calculate, and a simpler method might then be preferable The coefficient studied here has in most cases both properties mentioned above and can be used whenever its efficiency is not too small.

Let $(x_1, y_1) \cdots (x_n, y_n)$ be a sample from a two-dimensional population with cdf $F(x, y)$, and consider the two sample medians $\tilde{x}$ and $\tilde{y}$. The cdf $F(x, y)$ is assumed to have continuous marginal cdf's $F_1(x)$ and $F_2(y)$ in order that the probability of obtaining two equal x-values or two equal y-values in the sample will be zero. Let the x, y-plane be divided into four regions by the lines $x = \tilde{x}$ and $y = \tilde{y}$. It is then clear that some information about the correlation between x and y can be obtained from the number of sample points, say n_1, belonging to the first or third quadrants compared with the number, say n_2, belonging to the second or fourth quadrants.

Before going further we shall explain what is meant here by 'belong to'. If the sample size n is an even number the calculation of n_1 and n_2 is evident. If, however, n is an odd number one or two sample points must fall on the lines $x = \tilde{x}$ and $y = \tilde{y}$. In the first case this sample point shall not be counted. In the other case one point falls on each of the lines. Then one of the points shall be said to belong to the quadrant touched by both points, while the other shall

not be counted. It is easy to verify that both n_1 and n_2 by this method will be even numbers.

As a measure of correlation we define

$$(1) \qquad q' = \frac{n_1 - n_2}{n_1 + n_2} = \frac{2n_1}{n_1 + n_2} - 1 \qquad (-1 \leq q' \leq 1).$$

The definition of q' is not new [5] but as far as is known, its statistical properties have never been studied completely.

3. The asymptotic distribution. It is known [1] that the median in a sample from a one-dimensional distribution under certain conditions is a consistent estimate of the population median and asymptotically normally distributed. Although it seems possible to weaken the requirements in our case, we shall not do so. We require that

a) the population medians are uniquely defined (and assumed to equal zero),

b) the marginal distributions of $F(x, y)$ admit density functions $f_1(x)$ and $f_2(y)$.

c) $f_1(x)$, $f_2(y)$ and their first derivatives are continuous in some neighbourhood of the origin and

d) $f_1(0)$ and $f_2(0)$ are $\neq 0$.

In order to avoid trivial complications we shall assume here that the sample size $n = 2k + 1$.

Now define for every arbitrarily chosen point (x, y)

$$(2) \qquad \begin{aligned} a(x, y) &= P\{\xi > x, \eta > y\}, \\ b(x, y) &= P\{\xi \leq x, \eta > y\}, \\ c(x, y) &= P\{\xi \leq x, \eta \leq y\}, \\ d(x, y) &= P\{\xi > x, \eta \leq y\}, \end{aligned}$$

where the measure P refers to the cdf $F(x, y)$ and evidently

$$a + b + c + d = 1.$$

As the number of sample points belonging to the first and third quadrants around $(\tilde{x}, \tilde{y})$ must be equal, the probability of the combined event

$$\{n_1 = 2r; \tilde{x}\epsilon(x, x + dx), \tilde{y}\epsilon(y, y + dy)\}$$

is

$$(3) \qquad p_k(2r; x, y) = \frac{(2k + 1)!}{r!^2 \cdot (k - r)!^2} \cdot (ac)^r \cdot (bd)^{k-r} \cdot S,$$

where

$$(4) \quad S = \frac{r}{a} \cdot d_x a \cdot d_y a - \frac{k - r}{b} \cdot d_x b \cdot d_y b + \frac{r}{c} \cdot d_x c \cdot d_y c - \frac{k - r}{d} \cdot d_x d \cdot d_y d + dF.$$

Each of the first four terms of the expression (4) refers to a case in which two sample points determine $(\tilde{x}, \tilde{y})$, and the last term refers to a case in which $(\tilde{x}, \tilde{y})$ is determined by only one point. From (3) it follows that the probability of obtaining n_1 at most equal to $2R$ is

$$(5) \qquad P\{n_1 \leq 2R\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sum_{r=0}^{R} p_k(2r; x, y).$$

If we introduce the joint cdf $\Psi_k(x, y)$ of $\tilde{x}$ and $\tilde{y}$, (5) can be written

$$(6) \qquad P\{n_1 \leq 2R\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\Psi_k(x, y) \frac{\sum_{r=0}^{R} p_k(2r; x, y)}{\sum_{r=0}^{k} p_k(2r, x, y)},$$

as

$$d\Psi_k(x, y) = \sum_{r=0}^{k} p_k(2r; x, y).$$

Clearly the integrand in (6) is ≤ 1 everywhere it exists. In the points (x, y) where the denominator is equal to zero the integrand is undefined, but as the measure (Ψ) of the set of such points is zero, we need not have any trouble with them.

Under the conditions a)–d) $\tilde{x}$ and $\tilde{y}$ converge in probability to zero; that is

$$\lim_{k\to\infty} \Psi_k(x, y) = \begin{cases} 1 \text{ for } \{x \geq 0, y \geq 0\}, \\ 0 \text{ otherwise.} \end{cases}$$

Thus, when k and R tend to infinity such that $\frac{R}{k} \to$ const, (6) becomes

$$(7) \qquad \lim P\{n_1 \leq 2R\} = \lim \frac{\sum_{r=0}^{R} p_k(2r; 0, 0)}{\sum_{r=0}^{k} p_k(2r, 0, 0)}.$$

According to (3)

$$(8) \qquad p_k(2r; 0, 0) = \frac{(2k+1)!}{r!^2 \cdot (k-r)!^2} \cdot (a_0c_0)^r \cdot (b_0d_0)^{k-r} \cdot S_0,$$

where the subscripts indicate the value at the point (0, 0). Because of (2),

$$c_0 = a_0, \qquad d_0 = b_0 \quad \text{and} \quad a_0 + b_0 = \tfrac{1}{2},$$

and the two parts of (8) are for large k

$$\frac{(2k+1)!}{r!^2 \cdot (k-r)!^2} \cdot a_0^{2r} \cdot b_0^{2(k-r)} \sim \frac{1}{2\pi a_0 b_0 \sqrt{2\pi k}} \cdot e^{-((r-2ka_0)^2/4ka_0b_0)}$$

and

$$S_0 \sim 2k\left[\left(\frac{\partial a}{\partial x}\right)_0\left(\frac{\partial a}{\partial y}\right)_0 - \left(\frac{\partial b}{\partial x}\right)_0\left(\frac{\partial b}{\partial y}\right)_0 + \left(\frac{\partial c}{\partial x}\right)_0\left(\frac{\partial c}{\partial y}\right)_0 - \left(\frac{\partial d}{\partial x}\right)_0\left(\frac{\partial d}{\partial y}\right)_0\right] dx\, dy.$$

The first of these expressions follows from the usual application of Stirling's approximation formula and we omit all details here.

Hence, after the introduction of

$$r = 2ka_0 + t\sqrt{2ka_0b_0}\,,$$
$$R = 2ka_0 + T\sqrt{2ka_0b_0}\,,$$

the expression (7) is transformed to

$$\lim P\left\{\frac{n_1 - 4ka_0}{\sqrt{8ka_0b_0}} \leq T\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{T} e^{-\frac{1}{2}t^2}\, dt. \tag{9}$$

From (9) it follows that n_1 is asymptotically normally distributed with mean $4ka_0$ and standard deviation $\sqrt{8ka_0b_0}$. Thus

$$q' = \frac{2n_1}{2k} - 1 = \frac{n_1}{k} - 1$$

is asymptotically normally distributed with mean $4a_0 - 1$ and standard deviation $2\sqrt{a_0(1 - 2a_0)/k}$.

4. Properties as an estimator. Suppose we measure the correlation between x and y by

$$q = 2\left[\int_{-\infty}^{0}\int_{-\infty}^{0} dF + \int_0^{\infty}\int_0^{\infty} dF\right] - 1 = 4a_0 - 1, \tag{10}$$

where, as before, (0, 0) are the coordinates of the population medians. Then q has the desired property of being equal to zero in the case of independence and equal to ± 1 in the case of linear relationship between x and y.

According to (9) q' is a consistent estimate of q when the conditions a)–d) are fulfilled. Furthermore, as the standard deviation of q' is, to a first approximation, independent of quantities other than q, it is possible to construct approximate confidence limits for q for large sample sizes. This is done in the following way. In terms of n and q we have, according to the last paragraph of section 3 and (10),

$$Eq' \sim q,$$
$$\sigma(q') \sim \sqrt{\frac{1 - q^2}{n}}.$$

Let $\Phi(x)$ be a standardized normal cdf and λ_1 and λ_2 two numbers such that

$\Phi(\lambda_2) - \Phi(\lambda_1) = 1 - \alpha$. According to (9) we then have

$$P\left\{\lambda_1 < \frac{q' - q}{\sqrt{1 - q^2}} \cdot \sqrt{n} < \lambda_2\right\} \sim 1 - \alpha, \tag{11}$$

which gives the desired result.

If we let $\lambda_2 = -\lambda_1 = \lambda$ and solve the inequality in (11) for q, the following symmetrical confidence interval is obtained

$$q' - \frac{\lambda}{n}\sqrt{\lambda^2 + n(1 - q'^2)} < q < q' + \frac{\lambda}{n}\sqrt{\lambda^2 + n(1 - q'^2)},$$

where we have used that $\lambda^2 \ll n$.

5. The normal case. If x and y are normally distributed with correlation coefficient ρ, we have

$$q = \frac{2}{\pi}\arcsin \rho. \tag{12}$$

This expression is the same as the mean of Esscher-Kendall's rank correlation coefficient τ [2, 4]. Hence, in the normal case q' and τ estimate the same quantity. The coefficient q' has, however, a much smaller efficiency. The asymptotic efficiency of q' relative to the afore mentioned coefficient is

$$\frac{\sigma^2(\tau)}{\sigma^2(q')} \sim \frac{\frac{4}{n}\cdot\left[\frac{1}{9} - \left(\frac{2}{\pi}\arcsin\frac{\rho}{2}\right)^2\right]}{\frac{1}{n}\cdot\left[1 - \left(\frac{2}{\pi}\arcsin\rho\right)^2\right]} = \frac{4}{9}$$

for $\rho = 0$.

6. Tests of independence based on q'. In testing independence between x and y it is in practice more convenient to use critical regions based on n_1 instead of q'. Since, under the null hypothesis, the measure of a critical region is independent of $F(x, y)$ ($F_1(x)$ and $F_2(y)$ are assumed to be continuous), any test based on n_1 is non-parametric. We have made exact calculations of the q'-distribution for sample sizes n up to 50. For larger sample sizes the normal approximation for n_1 does not seem to entail errors of practical importance.

To derive the exact distribution of n_1 under the null hypothesis we suppose that n equals $2k$. The probability that any k sample points shall have smaller x-values than the other k points is

$$\binom{2k}{k}^{-1}.$$

Hence, since any arrangement of the sample points according to their x-values does not affect the distribution of the y-values,

$$P\{n_1 = 2r\} = \frac{\binom{k}{r}^2}{\binom{2k}{k}}. \tag{13}$$

If $n = 2k + 1$ it is easily verified that the probability (13) remains unchanged, if we use the procedure in calculating n_1 and n_2 proposed in Section 2. This is, in fact, the main reason for the proposal.

Table of $P\{|n_1 - k| \geq r\}$

r \ $2k$	4	8	12	16	20	24	28	32	36	40	44	48
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	.333	.486	.567	.619	.656	.681	.706	.724	.740	.752	.764	.773
4		.029	.080	.132	.179	.220	.257	.289	.318	.343	.366	.387
6			.0022	.010	.023	.039	.057	.076	.094	.113	.131	.148
8				.0002	.0011	.0033	.0070	.012	.018	.026	.034	.042
10						.0001	.0004	.0011	.0022	.0038	.0060	.0087
12									.0002	.0004	.0007	.0012
14											.0001	.0001

r \ $2k$	6	10	14	18	22	26	30	34	38	42	46	50
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	.100	.206	.286	.347	.395	.434	.466	.494	.517	.538	.556	.572
5		.0079	.029	.057	.086	.115	.143	.169	.194	.217	.238	.258
7			.0006	.0034	.0089	.017	.027	.038	.050	.063	.076	.089
9					.0003	.0012	.0028	.0053	.0086	.013	.017	.023
11							.0001	.0004	.0009	.0017	.0028	.0042
13									.0001	.0001	.0003	.0005
15												

$2k$ is the largest even number contained in the sample size

The distribution of n_1 is symmetric about $n_1 = k$ with the variance

$$\frac{k^2}{2k - 1}.$$

Thus, in testing independence we can for large sample sizes use

$$\frac{n_1 - k}{\sqrt{k}} \cdot \sqrt{2k - 1}$$

as an approximately normally distributed random variable with mean zero and unit s.d.

7. The asymptotic efficiency of the q'-test. In the case that x and y are normally distributed with the correlation coefficient ρ, it is possible, but rather tedious, to calculate the power function of the q'-test. We will, therefore, restrict ourselves to considering only the asymptotic behavior of the power function.

Consider tests of independence ($\rho = 0$) against one-sided alternatives $\rho > 0$. Let $L_m^{(1)}(\rho)$ be the power function of the q'-test for the sample size m and $L_n^{(2)}(\rho)$ be the power function of the test based on the correlation coefficient r in a sample of size n We assume that all tests have the same size, i.e.

$$L_m^{(1)}(0) = L_n^{(2)}(0) = \alpha \tag{14}$$

for all m and n. We shall say that the q'-test has the asymptotic efficiency ϵ if

$$\lim_{n\to\infty} \frac{\left(\frac{\partial L^{(1)}}{\partial \rho}\right)_{\rho=0}}{\left(\frac{\partial L^{(2)}}{\partial \rho}\right)_{\rho=0}} = 1 \tag{15}$$

when

$$m = \frac{n}{\epsilon}.$$

This means that the sample size in using the r-test need only be $100\epsilon\%$ of that in using the q'-test, in order to get the same derivative of the power functions at $\rho = 0$ (for large sample sizes). Since the definition of ϵ only concerns the behavior in the neighborhood of $\rho = 0$, it might perhaps be more correct to call ϵ the asymptotic local efficiency

In order to calculate ϵ we define two sequences $\{q_m\}$ and $\{r_n\}$ such that $\{q' > q_m\}$ and $\{r > r_n\}$ are tests with the afore mentioned properties. According to (9) and (10) q' is asymptotically normally distributed with mean q and s.d. $\sqrt{(1 - q^2)/m}$. Furthermore, r is asymptotically normally distributed with mean ρ and s.d. $(1 - \rho^2)/\sqrt{n}$. Hence,

$$1 - L_m^{(1)}(\rho) = P\{q' \le q_m \mid \rho\} \sim \Phi\left[\frac{q_m - q}{\sqrt{1 - q^2}} \sqrt{m}\right],$$

$$1 - L_n^{(2)}(\rho) = P\{r \le r_n \mid \rho\} \sim \Phi\left[\frac{r_n - \rho}{1 - \rho^2} \cdot \sqrt{n}\right],$$

from which it follows

$$\begin{aligned} \left(\frac{\partial L^{(1)}}{\partial \rho}\right)_0 &\sim \Phi'(q_m \cdot \sqrt{m}) \cdot \left(\frac{dq}{d\rho}\right)_0 \sqrt{m}, \\ \left(\frac{\partial L^{(2)}}{\partial \rho}\right)_0 &\sim \Phi'(r_n \cdot \sqrt{n}) \cdot \sqrt{n}. \end{aligned} \tag{16}$$

According to (14) we have

$$\lim_{m\to\infty} q_m \cdot \sqrt{m} = \lim_{n\to\infty} r_n \cdot \sqrt{n} = \Phi^{-1}(1 - \alpha).$$

Thus we conclude

$$\lim_{m,n\to\infty} \frac{\left(\frac{\partial L^{(1)}}{\partial \rho}\right)_0}{\left(\frac{\partial L^{(2)}}{\partial \rho}\right)_0} = \lim_{m,n\to\infty} \left(\frac{dq}{d\rho}\right)_0 \cdot \sqrt{\frac{m}{n}}. \tag{17}$$

Clearly (17) is equal to 1 if

$$\frac{n}{m} = \left(\frac{dq}{d\rho}\right)_0^2.$$

Hence, according to (12) and (15)

$$\epsilon = \left(\frac{2}{\pi}\right)^2.$$

In other words, the asymptotic efficiency of the q'-test is about 41%.

8. Concluding remarks. An interesting similarity exists between the q'-test of independence and a test of equal location parameters in two distributions, constructed in the following way. Suppose that two samples of equal size, say k, are drawn independently from two distributions. Compute the number of individuals, say r, in the first sample, falling short of the median of the pooled samples. Then the distribution of $2r$ under the null hypothesis is the same as that of n_1 in the q'-test for sample size $2k$ (or $2k + 1$). The test based on r was discussed by F. Mosteller [7].

Another similarity is between the q'-test and a special case of the exact test of independence in a 2 x 2 table [8]. If in such a table the marginals happen to be cut at the 50% points the two test procedures become identical.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[2] F. Esscher, "On a method of determining correlation from the ranks of a variate", *Skandinavisk Aktuarietidskrift*, Vol. 7 (1924), p. 201.
[3] W. Hoeffding, "A non-parametric test of independence", *Annals of Math. Stat.*, Vol. 19 (1948), p. 546
[4] M. G. Kendall, "A new measure of rank correlation", *Biometrika*, Vol. 30 (1938), p. 81.
[5] F. Mosteller, "On some useful 'inefficient' statistics", *Annals of Math. Stat.*, Vol. 17 (1946), p. 377.
[6] C. Spearman, "The proof and measurement of association between two things", *Am. Jour. of Psych.*, Vol. 15 (1904), p. 88.
[7] F. Mosteller, "On some useful 'inefficient' statistics", unpublished thesis, Princeton University, 1946.
[8] R. A. Fisher, *Statistical Methods for Research Workers*, 8th Ed, Stechert & Co., 1941.

SOME TWO SAMPLE TESTS

By Douglas G. Chapman[1]

University of Washington

1. Introduction and summary. Stein [4] has exhibited a double sampling procedure to test hypotheses concerning the mean of normal variables with power independent of the unknown variances. This procedure is here adapted to test hypotheses concerning the ratio of means of two normal populations, also with power independent of the unknown variances. The use of a two sample procedure in a regression problem is also considered.

Let $\{X_{ij}\}$ $(i = 1, 2)$ $(j = 1, 2, 3, \cdots)$ be independent random variables distributed according to $N(m_i, \sigma_i)$: all parameters are assumed to be unknown.

Defining k by the equation

$$m_1 = km_2 \tag{1}$$

we wish to test the hypothesis H that k has a specified value k_0.

If $k_0 = 1$ the hypothesis H reduces to a classical problem, often referred to in the literature as the Behrens-Fisher-problem (cf. Scheffé [3] for a bibliography). At the present time it is still an open question whether it is possible (or desirable) to find a non-trivial single sample test for H with the *size* of the critical region independent of σ_1 and σ_2. In any case it is a simple extension of the result of Dantzig [1] (cf. also Stein [4]) to show that no non-trivial single sample test exists whose *power* is independent of σ_1 and σ_2.

On the other hand the case $k_0 \neq 1$ may be expected to occur frequently in fields of application where a choice must be made between different products, methods of experimentation etc. which involve different costs. The statistician must make a choice on the basis of results relative to the ratio of costs involved. Nevertheless this problem appears to have received little attention in the literature.

In general tests based on a two-sample procedure may not be as "efficient" in the sense of Wald [5] as a strict sequential procedure. On the other hand the two sample procedure reduces the number of decisions to be made by the experimenter and it will, in certain fields, simplify the experimental procedure.

2. The two sample procedure. Stein's double sampling procedure (which may be denoted procedure S) to test a hypothesis concerning the mean of a normal population consists briefly in the following steps:

(a) Choose "a priori" a positive number z and a preliminary sample size n.

(b) Take n independent observations $x_1, \cdots, x_n$ of the random variable X

[1] This research was carried out while the author was at the University of California, Berkeley, and was supported in part by the Office of Naval Research.

which is assumed to be distributed according to $N(m, \sigma^2)$ with unknown mean m and unknown variance σ^2, and calculate

$$(2) \qquad u^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n - 1}.$$

(c) Let $N = \max\left(\left[\frac{u^2}{z}\right] + 1,\ n + 1\right)$ where $[r]$ = largest integer $\leq r$

(d) Take $N - n$ more independent observations of X and choose a set of constants $a_1, \cdots a_N$ such that

$$(3) \qquad \text{(i)} \sum_{i=1}^{N} a_i = 1, \qquad \text{(ii)}\ a_1 = a_2 = \cdots = a_n, \qquad \text{(iii)} \sum_{i=1}^{N} a_i^2 = \frac{z}{u^2}.$$

(e) Then $\dfrac{\sum_{i=1}^{N} a_i x_i - m}{\sqrt{z}}$ has Student's t-distribution with $n - 1$ degrees of freedom.

Stein further showed that the procedure may be modified to some advantage in problems dealing with a single population. This modification is not applicable in the problems under consideration here.

There remains to be discussed briefly the choice of n, z and the a's. The preliminary sample size n may be determined by other considerations or it may be chosen as part of the design of the experiment. Hodges [2] has shown that the expected value of the total sample size N and the power of the test both depend on the choice of n and he has discussed the optimum choice of n with respect to the modified procedure of Stein. In general this optimum choice of n depends upon prior knowledge concerning the variance.

The power of the test will depend upon z: some considerations concerning the choice of z will be dealt with after discussing the tables upon which the two sample tests are based.

The arbitrariness involved in choosing the a's may be eliminated by placing the additional requirement that

$$(4) \qquad a_{n+1} = a_{n+2} = \cdots = a_N = b \quad \text{(say)}.$$

Letting $a_1 = a_2 = \cdots = a_n = a$ it is elementary to solve for a and b explicitly viz.,

$$(5) \qquad \begin{aligned} na + (N - n)b &= 1, \\ na^2 + (N - n)b^2 &= \frac{z}{u^2}. \end{aligned}$$

The solutions are

$$(6) \qquad b = \frac{1}{N}\left(1 + \sqrt{\frac{n(Nz - u^2)}{(N - n)u^2}}\right),$$

$$(7) \qquad a = \frac{1 - (N - n)b}{n}.$$

3. Test for H. The steps involved in testing the hypothesis H are

(a) Choose the preliminary sample size n, and positive numbers z_1, z_2 subject to the restriction

$$\frac{z_1}{z_2} = k_0^2. \tag{8}$$

(b) Carry out procedure S with the same n for each population, determining two statistics T_1, T_2, i e

$$T_i = \frac{\sum_{j=1}^{N_i} a_{ij} x_{ij}}{\sqrt{z_i}} \qquad (i = 1, 2). \tag{9}$$

Then $T_1 - T_2$ has, under the hypothesis tested, the distribution of the difference of two independent Student variables.

If s denotes the difference of two independent random variables t_1 and t_2 each distributed according to Student's t-distribution with $n - 1$ degrees of freedom and if s_0 is defined by the equation

$$P(|s| > s_0) = \alpha,$$

then a test of size α is given by the rule: H is rejected if $|T_1 - T_2| > s_0$

4. The distribution of differences of Student variables. The distribution of s is easily found by the method of characteristic functions, in case n is even.

Let $m = n - 1$ and to simplify slightly put

$$y_i = \frac{t_i}{\sqrt{m}} \qquad (i = 1, 2). \tag{10}$$

Then the density function of y_i is

$$f(y) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{m}{2}\right)} \frac{1}{(1 + y^2)^{(m+1)/2}} \tag{11}$$

and its characteristic function

$$\varphi_y(t) = \int_{-\infty}^{+\infty} e^{iyt} f(y)\, dy \tag{12}$$

$$= \frac{\sqrt{\pi}}{\Gamma\left(\frac{m}{2}\right)} \frac{e^{-|t|}}{2^{m-1}} \left(\sum_{r=0}^{(m-1)/2} \frac{\left(\frac{m-1}{2} + r\right)!}{m!\left(\frac{m-1}{2} - r\right)!} [2(|t|)]^{(m-1)/2-r} \right). \tag{13}$$

Formula (13) may be obtained by contour integration, it is, however, a standard formula in connection with Bessel functions of the second kind of purely imaginary argument (cf. Watson [6], pp. 80, 185–188).

While it is not possible to obtain a simple general expression for

$$f(w) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iwt}[\varphi_y(t)]^2\,dt, \tag{14}$$

the density function of $w = \frac{s}{\sqrt{m}}$, this integral may be evaluated for $m = 1, 3, 5$ etc. and furthermore the density function of s may be integrated in a closed form for such values of m, and consequently tabulated fairly easily.

In case n is odd it is possible to express $\varphi_y(t)$ in terms of Bessel functions but the Bessel functions obtained are not expressible in a closed form. While the problem may be attacked directly by numerical integration, it will generally be sufficient to interpolate in Table I where necessary, for such values of n.

Table I gives the distribution of s for $n = 2, 4, 6, 8, 10, 12$. For larger values of n it may be sufficiently accurate to use the normal approximation to the distribution of s. In virtue of the asymptotic normality of the t-distribution s will be distributed approximately normally with mean zero and variance $\frac{2(n-1)}{n-3}$ for n sufficiently large.

5. Power of the test. Writing

$$\Delta = \frac{m_1}{\sqrt{z_1}} - \frac{m_2}{\sqrt{z_2}} \quad \text{and} \quad T = T_1 - T_2 \tag{15}$$

it is seen that $T = s + \Delta$ and hence

$$P(H \text{ is rejected}) = P(|T| > s_0) = P(s < -s_0 - \Delta) + P(s > s_0 - \Delta). \tag{16}$$

Since

$$\Delta = \frac{m_2}{\sqrt{z_2}}\left(\frac{k}{k_0} - 1\right)$$

equation (16) may be used as a guide in choosing z_2 so that a certain minimum power is attained; the presence of the nuisance parameter m_2 makes impossible the determination of z_2 so as to give exactly some preassigned power.

Since s is distributed independently of σ_1, σ_2, it follows that the power of the test is independent of these parameters. Using the addition formula to express the frequency function of s in terms of the frequency function of Students' t-distribution, it may be shown that $f(s)$ in unimodal and symmetrical about $s = 0$. Hence the test is unbiased. It also follows from (16) that if z_2 is made to approach zero the probability of rejecting H when it is false tends to 1: i.e. the test is consistent.

It may be observed that tests for the one-sided hypotheses

$$\frac{m_1}{m_2} \geq k \quad \text{or} \quad \frac{m_1}{m_2} \leq k$$

may easily be formulated. Table II provides a table useful for such tests also, at half the indicated significance levels.

TABLE I

Distribution of s: difference of two independent student-variables with n − 1 degrees of freedom
The value tabled is $P(0 \leq s \leq s_o)$

s_o \ n	2	4	6	8	10	12	Normal Approximation for $n = 12$
0.50	0 0780	0.1014	0.1222	0.1265	0 1290	0.1306	0 1254
1.00	.1476	.1922	.2311	.2392	.2438	.2467	2388
1.50	.2048	2660	.3185	.3290	3349	3386	3313
2.00	2500	.3243	.3825	3939	.4002	.4041	3996
2.50	.2852	.3620	4260	.4364	.4415	.4465	4451
3.00	.3128	.3903	4542	.4637	4687	4724	.4725
3.50	.3348	.4104	4726	.4796	4834	4856	.4874
4.00	.3524	4247	.4825	.4884	.4914	4929	.4947
4.50	.3669	.4352	.4890	4936	4956	.4966	.4980
5 00	.3789	.4431	.4930	.4964	.4977		
5.50	.3890	.4491	4955	4980	.4988		
6 00	3976	4539	4970	.4988			
6 50	.4050	.4578	4980				
7.00	.4114	.4611	.4986				
7.50	.4170	.4638					
8.00	.4220	4661					
10.00	.4372	.4730					
12.00	.4474	4774					
21.00	.4698	4870					
30.00	.4788	.4908					
50.00	.4873						
100 00	.4936						

TABLE II

The 5% and 1% significance points of the distribution of s
The value tabled is s_o

Significance Level \ n	2	4	6	8	10	12	Normal Approximation for $n = 12$
$P(\vert s \vert \geq s_o) = .05$	25 41	10.82	3.62	3.34	3 18	3 10	3 06
$P(\vert s \vert \geq s_o) = .01$	127.3	36.8	5 38	4 72	4 42	4 26	4 03

6. A regression problem. We consider the problem where x_i are values of a sure variable, Y_i are independent random variables with

$$E(Y_i) = a + bx_i \tag{17}$$

and σ_{Y_i} is unknown. It is desired to estimate a and b and to test the hypothesis $b = b_0$.

The usual procedure is to assume $\sigma^2_{Y_i}$ constant, and use the Markov theorem (i.e. the standard least squares formulae). In this way unbiased estimates of a and b are obtained, whether or not this assumption is fulfilled. However the usual significance test for b is not valid if this assumption (plus normality of the Y's) is not fulfilled.

The two sample procedure leads to a valid test of the hypothesis $b = b_0$, with power independent of the unknown variance. Since linearity of the expected value of Y on x is assumed, the optimum procedure is to observe Y for only two values of x, at opposite ends of the range. Let these points be x_1, x_2. For these values of x, procedure S may be used (choosing $z_1 = z_2$) to determine T_1, T_2 where $T_i - (a + bx_i)/\sqrt{z}$ has Student's t-distribution with $n - 1$ degrees of freedom.

Then the following estimates of a, b are unbiased, for $n \geq 3$,

$$\hat{b} = \left(\frac{T_2 - T_1}{x_2 - x_1}\right)\sqrt{z}, \tag{18}$$

$$\hat{a} = \left(\frac{x_2 T_1 - x_1 T_2}{x_2 - x_1}\right)\sqrt{z}. \tag{19}$$

To test the hypothesis $H_1 : b = b_0$ it is necessary only to calculate the statistic $\zeta = [(T_1 - T_2)\sqrt{z} - b_0(x_1 - x_2)]/\sqrt{z}$ and reject H_1, at the α level of significance if $|\zeta| > s_0$, where s_0 was defined above (Section 3).

It is seen that if b' is the true value of b, then the power of the test is a function of $(b' - b_0)(x_1 - x_2)/\sqrt{z}$ and z may be determined to obtain any prescribed power desired. It is also immediate that the power of the test is independent of σ_{Y_i}.

The author wishes to express thanks to the members of the computing staff of the Statistical Laboratory, University of California, Mrs. E. Putz, Miss J. Linton, and Mr. J. Blum, for assistance in preparing Tables I and II.[2]

REFERENCES

[1] GEORGE B. DANTZIG, "On the non-existence of tests of 'Student's' hypothesis having power functions independent of σ," *Annals of Math. Stat.*, Vol. 11 (1940), p. 186.

[2] JOSEPH L. HODGES, JR., "The selection of initial sample size in the Stein two sample procedure", unpublished dissertation, University of California, Berkeley, 1948.

[3] HENRY SCHEFFÉ, "On solutions of the Behrens-Fisher Problem based on the t-distribution", *Annals of Math. Stat*, Vol. 14 (1943), p. 35.

[4] CHARLES STEIN, "A two sample test for a linear hypothesis whose power is independent of the variance", *Annals of Math. Stat.*, Vol. 16 (1945), p. 243.

[5] ABRAHAM WALD, *Sequential Analysis*, John Wiley and Sons, Inc., 1947.

[6] G. N. WATSON, *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, 1944.

[2] It has been pointed out to the writer that percent points of linear combinations of two independent Student t's are given in Table VI (by P. V. Sukatme) in R. A FISHER AND F. YATES, *Statistical Tables for Biological, Medical and Agricultural Research*, Oliver and Boyd, Edinburgh, 1943 (added in page proof).

NOTES

This section is devoted to brief research and expository articles and other short items.

TRANSFORMATIONS RELATED TO THE ANGULAR AND THE SQUARE ROOT

By Murray F. Freeman and John W. Tukey[1]

Princeton University

1. Summary. The use of transformations to stabilize the variance of binomial or Poisson data is familiar (Anscombe [1], Bartlett [2, 3], Curtiss [4], Eisenhart [5]). The comparison of transformed binomial or Poisson data with percentage points of the normal distribution to make approximate significance tests or to set approximate confidence intervals is less familiar. Mosteller and Tukey [6] have recently made a graphical application of a transformation related to the square-root transformation for such purposes, where the use of "binomial probability paper" avoids all computation We report here on an empirical study of a number of approximations, some intended for significance and confidence work and others for variance stabilization.

For significance testing and the setting of confidence limits, we should like to use the normal deviate K exceeded with the same probability as the number of successes x from n in a binomial distribution with expectation np, which is defined by

$$\frac{1}{2\pi}\int_{-\infty}^{K} e^{-\frac{1}{2}t^2}\,dt = \text{Prob}\,\{x \leq k \mid \text{binomial}, n, p\}.$$

The most useful approximations to K that we can propose here are N (very simple), N^+ (accurate near the usual percentage points), and N^{**} (quite accurate generally), where

$$N = 2\left(\sqrt{(k+1)q} - \sqrt{(n-k)p}\right).$$

(This is the approximation used with binomial probability paper.)

$$N^+ = N + \frac{N + 2p - 1}{12\sqrt{E}}, \qquad E = \text{lesser of } np \text{ and } nq,$$

$$N^* = N + \frac{(N-2)(N+2)}{12}\left(\frac{1}{\sqrt{np+1}} - \frac{1}{\sqrt{nq+1}}\right),$$

$$N^{**} = N^* + \frac{N^* + 2p - 1}{12\sqrt{E}}. \qquad E = \text{lesser of } np \text{ and } nq.$$

For variance stabilization, the averaged angular transformation

$$\sin^{-1}\sqrt{\frac{x}{n+1}} + \sin^{-1}\sqrt{\frac{x+1}{n+1}}$$

[1] Prepared in connection with research sponsored by the Office of Naval Research.

has variance within ±6% of

$$\frac{1}{n+\frac{1}{2}} \text{ (angles in radians)}, \qquad \frac{821}{n+\frac{1}{2}} \text{ (angles in degrees)},$$

for almost all cases where $np \geq 1$.

In the Poisson case, this simplifies to using

$$\sqrt{x} + \sqrt{x+1}$$

as having variance 1.

2. Significance testing. In addition to the approximations mentioned above, empirical study was also made of the following

$$L = \frac{x - np}{\sqrt{npq}},$$

$L^* = L$ modified by a term like that in N^*,

$$M = 2\sqrt{n+1}\left(\sin^{-1}\sqrt{\frac{k+1}{n+1}} - \sin^{-1}\sqrt{p}\right),$$

$M^* = M$ modified by a term like that in N^*.

Taking an upper limit of 2.5 or 3.5 on $|K|$ and a lower limit of 0.01, 1, or 4 on np, the greatest observed errors of the approximations were smallest for N^{**}, N^* and M^* and largest for the direct approximations L and L^*. This was true for all six choices of region.

If we exclude the cases $k = 0$ and $k = n$, where the desired probability can be calculated directly, the largest observed errors in the substantial number of cases computed, which are probably representative of the regions where the approximations are worst, were as follows:

$\|K\|$	$E = np$	Largest observed error of N^{**}	M^*	N^*	N^+	N	M	L^*	L
≤ 2.5	≥ 4	.04	.07	.08	.14	.16	.17	.26	.35
	≥ 1	.04	.09	.13	.19	.20	.24	.35	.42
	≥ 0.01	.04	.20	.20	.19	.20	.65	.62	.80
≤ 3.5	≥ 4	.08	.07	.08	.19	.25	.25	.57	.63
	≥ 1	.11	.10	.17	.21	.38	.34	1.51	1.26
	≥ 0.01	.11	.51	.60	.21	.65	.65	5.88	3.42

Within the range of great interest, $|K| \leq 2.5$, that is $.0062 \leq$ probability $\leq .9938$, we have errors of less than 0.04 in N^{**} and less than 0.20 in N.

For $1.5 < |K| < 2.5$, the range of greatest interest, the average error of N^+ was less than 0.03 and the maximum was 0.08 (54 cases considered).

Thus, we can recommend

N as a simple and usually accurate transformation,

N^+ for rapid significance testing,

N^{**} for adequate accuracy at all levels.

Figure 1 shows the behavior of the various approximations in the case $n = 50$, $np = 5$. This is roughly typical.

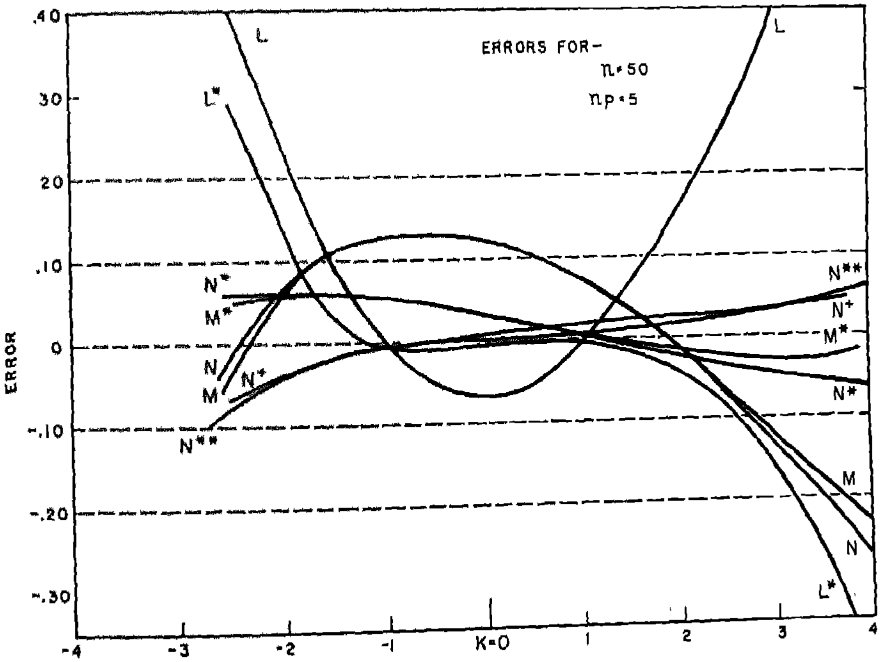


FIG. 1. Errors of approximation.

3. Variance stabilization. The various suggestions for stabilizing the variance of the Poisson are:

$\sqrt{x + 1/2}$, (Bartlett [2]),

$\sqrt{x + 3/8}$, (Anscombe [1]),

$\sqrt{x} + \sqrt{x + 1}$, (this paper).

Figure 2 shows the variance of the transformed variate as a function of the Poisson expectation. Clearly $\sqrt{x} + \sqrt{x + 1}$ is the best if small expectations are to be considered. The simplicity with which it can be read from a square-root table, and its unit variance, are also favorable factors.

When an approximation of a given form is to work over as large a range as

possible without the magnitude of its errors exceeding a certain limit, the optimum approximation is almost certain to involve errors of *both* signs. If $\pm 6\%$ variation in variance is permissable, $\sqrt{x} + \sqrt{x+1}$ is usable for expectations of unity or more. It is not surprising that Anscombe's approximation, obtained by eliminating the term in n^{-1}, and dominated by the term in n^{-2}, should only meet the $\pm 6\%$ tolerance for expectations of 2.2 or more.

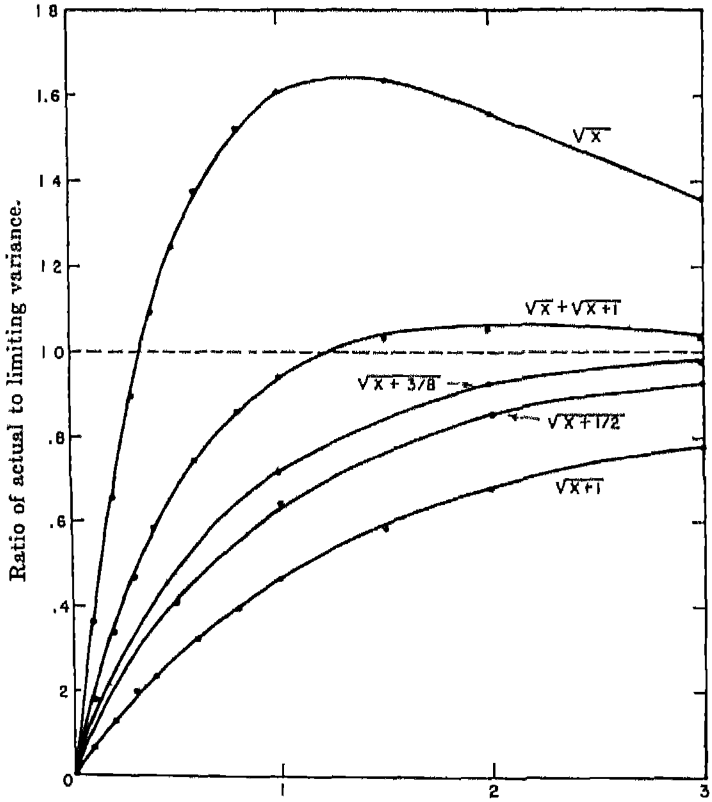


Fig. 2. Stabilization of Poisson variance.

4. Scope. Values of K, and with some occasional exceptions, of L, L^*, M^*, N, N^+, N^* and N^{**} were calculated for

$n = 2, 5, 10, 20, 100,$

$p = 1\%, 2\%, 5\%, 10\%, 20\%, 30\%, 40\%, 50\%,$

k giving $K < 4.5$,

and similar computations were made for the Poisson case with expectations

1/100, 1/50, 1/20, 1/10, 1/5, 1/2, 1, 2, 4, 8, 16, 32, 64.

These computations were made to only two decimal places, so that the final results may easily err by 1, 2, or 3 in the second decimal place.

A more complete discussion of the problem, the origin of the approximations, and tables showing a representative collection of actual values can be found in Memorandum Report 24 of the Statistical Research Group, Princeton University, which bears the same title as this note. Copies may be obtained from its Secretary, Box 708, Princeton, N. J.

REFERENCES

[1] F J Anscombe, "The transformation of Poisson, binomial, and negative binomial data", *Biometrika*, Vol 35 (1948), pp 246–254

[2] M S. Bartlett, "The square root transformation in the analysis of variance", *Jour Roy. Stat. Soc., Suppl.*, Vol. 3 (1936), pp 68–78

[3] M. S. Bartlett, "The use of transformations", *Biometrics*, Vol. 3 (1947), pp 39–51.

[4] J. H. Curtiss, "On transformations used in the analysis of variance", *Annals of Math. Stat.*, Vol. 14 (1943), pp. 107–122

[5] Churchill Eisenhart, "The assumptions underlying the analysis of variance", *Biometrics*, Vol. 3 (1947), pp 1–21.

[6] Frederick Mosteller and John W. Tukey, "The uses and usefulness of binomial probability paper", *Jour. Am Stat Assn* , Vol 44 (1949), pp 174–212

REMARK ON THE ARTICLE "ON A CLASS OF DISTRIBUTIONS THAT APPROACH THE NORMAL DISTRIBUTION FUNCTION" BY GEORGE B. DANTZIG[1]

By T. N. E. Greville

Federal Security Agency

In this interesting and valuable article, Dr. Dantzig showed that, under certain conditions, a sequence of frequency distributions connected by a linear recurrence formula converges to the normal distribution. Among several applications of his results which are discussed, the author mentions their relation to certain types of smoothing formulas, and has shown that if a linear smoothing formula and the data to which it is applied satisfy certain conditions, the iteration of the smoothing process produces a sequence of smoothed distributions which, upon normalization, approaches the normal frequency curve.

In a summary paragraph at the end of the article, it is stated that "successive application of one or many such linear formulas will usually smooth *any* set of values to the normal curve of error" The entire article was concerned with frequency distributions, and a careful reading makes it clear that the author intended the quoted statement to apply only to data in this form. However, its rather general wording seems to have led a number of readers to interpret it as being applicable to other types of data, such as time series, which frequently may not satisfy the conditions assumed Moreover, it is easy to overlook the

[1] *Annals of Math Stat.*, Vol. 10 (1939), pp. 247–253.

restrictions imposed on both the original data and the smoothing formula as they are stated only by implication, and not explicitly, even though they have the effect of excluding important classes of smoothing formulas, such as those commonly employed by actuaries.

The approach to the normal distribution is shown to depend on the vanishing of a certain limit denoted as Γ' which is a function of the moments of the original data and of a distribution in which the weights employed in the smoothing formula are interpreted as frequencies. At this point, objection may be taken to Dr. Dantzig's proof, since the smoothing formulas most frequently used contain negative weights. However, it has been shown elsewhere[2] that the occurrence of negative weights will not of itself prevent the sequence of smoothed distributions from approaching the normal curve. A somewhat more serious difficulty arises if, as is commonly the case, the smoothing formula has the property of reproducing polynomials of a specified degree. If the degree reproduced is two or more, this implies the vanishing of the second moment of the weight distribution, in which case the limit Γ' does not vanish. In fact, it has been shown by DeForest[3] and Schoenberg that the iteration of smoothing formulas which reproduce polynomials of higher degree gives rise to a sequence of limiting distributions which have the general appearance of the normal curve in the center portion and of a damped sine curve in the tails. This is, however, at best, a technical exception to Dantzig's statement, as one is still faced with his basic proposition that repeated application of a smoothing formula to a frequency distribution will cause the smoothed distribution to be dominated by the characteristics of the smoothing formula rather than those of the original data.

While he did not intend the statement to refer to data not in the form of a frequency distribution, some readers seem to have interpreted it as being of general application, and, for that reason, I should like to point out a few of the considerations involved in applying iterated smoothing to other types of data, such as, for example, a time series or the values of a mathematical function. The limit Γ', on whose vanishing Dantzig's theorem depends, involves the second and fourth moments of the original data (as well as of the weight distribution) and, therefore, can be computed only if these moments exist. For this it is necessary (but, of course, not sufficient) that the function being smoothed shall tend toward zero as the independent variable approaches positive or negative infinity.

In order to iterate a smoothing formula an infinite number of times, it is obviously necessary to have an infinite set of original values. Therefore, in smoothing, for example, a finite time series, one would have to make some assumption regarding the values of the series outside the range for which they

[2] I. J. Schoenberg, "Some analytical aspects of the problem of smoothing," *Courant Anniversary Volume*, Interscience Publishers, New York, 1948.

[3] H. H. Wolfenden, "On the development of formulae for graduation by linear compounding, with special reference to the work of Erastus L. DeForest," *Trans. Actuarial Soc. Am.*, Vol. 26 (1925), pp. 81–121

are actually available. Of course, if it were assumed that the values were zero outside this range, Dantzig's theorem would apply. However, under this assumption, infinite iteration of a smoothing formula would not be a rational procedure, as it would smooth each value to zero, and the incidental fact that the sequence of smoothed distributions, while approaching zero, also approach the form of a normal distribution, would not be a very valuable one. In this connection, an important distinction between time series and frequency data is that, in dealing with the former, one is interested in the magnitude of individual values as well as in the general form and shape of the distribution. In practice it might be preferable not to make any assumption about the values outside the given range but rather to employ special devices to obtain smoothed values near the ends of this range. In such a case, the smoothing process would be a function of the range (if not of the actual values) of the original data distribution. Such a process was not considered by Dantzig, and is clearly excluded by his definition of a linear smoothing formula, which requires that the formula be completely independent of the data to which it is applied.

The somewhat academic question of the effect of iteration of a smoothing formula on a function of infinite range for which the moments do not exist, is a difficult one, to which I cannot give a general answer. Schoenberg does not consider this problem, but merely gives the weight distribution to be applied to the original data in order to obtain the limiting smoothed distribution Two trivial examples may, however, serve to illustrate the nature of the considerations involved. If the original data are values of a polynomial of a specified degree, and if a smoothing formula which reproduces that degree is successively applied, it will of course continue indefinitely to reproduce the original values On the other hand, if the smoothing formula reproduces only polynomials of lower degree, a bias is introduced As a simple example, we may consider the case of smoothing the function $y = x^2$ by a formula consisting of three weights each equal to 1/3 to be applied to the given value and its two immediate neighbors. It is easily shown that the smoothed value is $x^2 + 1/3$, and the effect of successive application of this formula is to add 1/3 each time. Thus each smoothed value would tend toward infinity as the number of smoothings increases, however, the entire distribution would always remain a parabola of the same form as originally.

Finally, I should like to emphasize that, in common with Dr. Dantzig, I do not regard infinite repetition of the smoothing operation as a practical procedure, but consider it preferable to select, in the first instance, a smoothing formula which is likely to have the desired effect and then to perform the smoothing in a single step. In this way, one is more likely to secure the result desired without losing sight of important characteristics of the original data.

INDEPENDENCE OF QUADRATIC FORMS IN NORMALLY CORRELATED VARIABLES[1]

By Yukiyosi Kawada

Tokyo University of Literature and Science

The problem to give a necessary and sufficient condition that two quadratic forms in normally correlated variables are independent was treated by many authors [1], [2], [3], [4], [5]. We shall give here also a solution of this problem, which may be a generalization of that given by B. Matérn [6] for nonnegative quadratic forms to the general case.

Theorem 1. *If two quadratic forms*

$$Q_1 = \sum_{ij=1}^{n} a_{ij}x_ix_j, \quad Q_2 = \sum_{ij=1}^{n} b_{ij}x_ix_j \tag{1}$$

in normally correlated variables $x_1, \cdots, x_n$ *with zero means and with the variance matrix* I *satisfy the following four conditions*

$$F_{ij} = E(Q_1^iQ_2^j) - E(Q_1^i)E(Q_2^j) = 0 \qquad (i, j = 1, 2), \tag{2}$$

then the relation

$$AB = 0 \qquad (A = (a_{ij}), B = (b_{ij})) \tag{3}$$

holds.

Corollary 1. *If* Q_1, Q_2 *in* (1) *satisfy the four conditions* (2), *then* Q_1 *and* Q_2 *are independent.*

Corollary 2. (Necessity portion of the theorem of Craig) *A necessary condition for the independence of* Q_1 *and* Q_2 *is* $AB = 0$. (The sufficiency was proved by Craig.)

Proof of Theorem 1. The proof is very simple. Using the values $E(x_k^i) = 0$, $(i = 1, 3, 5, 7)$, $E(x_k^2) = 1$, $E(x_k^4) = 3$, $E(x_k^6) = 15$, $E(x_k^8) = 105$ $(k = 1, \cdots, n)$, we have by a straightforward calculation[2] the following relations

(4) $F_{11} = 2Tr(AB)$,

(5) $F_{12} = 8Tr(AB^2) + 4Tr(AB)Tr(B)$,

(6) $F_{21} = 8Tr(A^2B) + 4Tr(AB)Tr(A)$,

(7) $F_{22} = 32Tr(A^2B^2) + 16Tr((AB)^2) + 16Tr(AB^2)Tr(A) + 16Tr(A^2B)Tr(B)$
$+ 8Tr(AB)Tr(A)Tr(B) + 8Tr(AB)^2$.

---

[1] Presented at the Chapel Hill meeting of the Institute of Mathematical Statistics and Biometric Society March 18, 1950.

[2] If we apply an orthogonal transformation on $(x_1, \cdots, x_n)$ so that A becomes a diagonal form, the calculation becomes simpler than with the general form. We may note here also the fact that we need not assume that $x_1, \cdots, x_n$ are normally correlated, but we use only the values of $E(x_k^i)$ $(i = 1, \cdots, 8)$ for our proof.

Put $C = AB$. Let C' be the transposed matrix of C. We have from (2), (4)–(7)

$$2Tr(A^2B^2) + Tr((AB)^2) = 2Tr(CC') + Tr(C^2) = 0. \tag{8}$$

The left side of (8) is equal to $\sum_{i,j=1}^{n} (c_{ij}^2 + c_{ij}c_{ji} + c_{ji}^2)$, which is positive unless all $c_{ij} = 0$ $(i, j = 1, \cdots, n)$. Hence we have $C = AB = 0$, q.e.d

Corollary 1 follows from Theorem 1 and the theorem of Craig. Corollary 2 results from observing that independence of Q_1 and Q_2 implies (2).

B. Matérn proved, that if A, B are nonnegative, then $AB = 0$ follows from a unique condition $F_{11} = 2Tr(AB) = 0$. If only one of the matrices A, B is assumed to be nonnegative, we have

THEOREM 2. *Let A be nonnegative Then from two conditions $F_{11} = 0$, $F_{12} = 0$ in (2) follows the relation $AB = 0$*

PROOF. From (4), (5) follows $Tr(AB^2) = 0$. Since A is nonnegative, we can choose a real symmetric matrix A_0 such that $A = A_0^2$ Put $C_0 = A_0B$. Then we have $Tr(AB^2) = Tr(C_0C_0') = 0$ and from this follows $C_0 = 0$. Hence we have also $AB = A_0C_0 = 0$, q.e.d

REFERENCES

[1] A. T. CRAIG, "Note on the independence of certain quadratic forms", *Annals of Math. Stat*, Vol. 14 (1943), pp 195–197.

[2] H. HOTELLING, "Note on a matric theorem of A T. Craig", *Annals of Math Stat*, Vol 15 (1944), pp. 427–429

[3] H. SAKAMOTO, "On the independence of two statistics", *Research Memoirs of Inst of Stat. Math*, Tokyo, Vol 1 (1944), pp 1–25 (in Japanese).

[4] K. MATUSITA, "Note on the independence of certain statistics", *Annals of Inst of Stat. Math.*, Tokyo, Vol 1 (1949), pp 79–82

[5] J. OGAWA, "On the independence of bilinear and quadratic forms of a random sample from a normal population", *Annals of Inst of Stat. Math.*, Tokyo, Vol 1 (1949), pp. 83–108.

[6] B. MATÉRN, "Independence of non-negative quadratic forms in normally correlated variables", *Annals of Math. Stat*, Vol. 20 (1949), pp 119–120

ERRATA TO "CONTROL CHART FOR LARGEST AND SMALLEST VALUES"

BY JOHN M. HOWELL

Los Angeles City College

In the paper cited in the title (*Annals of Math. Stat.*, Vol. 20 (1949), p 306), there are some numerical errors in Table I Values of $d_2/2$ and d_4 are given by H. J. Godwin in "Some Low Moments of Order Statistics" in the same issue

of the *Annals*. These values are more accurate than those heretofore available. A corrected Table I based on these values is as follows:

n	d_2	d_3	A_2	A_3	A_4	n
2	1.1284	.8256	1.8800	2.6951	3.0411	2
3	1.6926	.7480	1.0233	1.8258	3.0902	3
4	2.0588	.7012	.7286	1.5218	3.1330	4
5	2.3259	.6690	.5768	1.3629	3.1699	5
6	2.5344	.6449	.4832	1.2634	3.2020	6
7	2.7043	.6260	.4193	1.1945	3.2303	7
8	2.8472	.6107	.3725	1.1434	3.2556	8
9	2.9700	.5978	.3367	1.1038	3.2784	9
10	3.0775	.5868	.3083	1.0720	3.2992	10

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Berkeley meeting of the Institute, August 5, 1950)

1. Sampling from Populations with Overlapping Clusters. Z. W. Birnbaum, University of Washington, Seattle.

In cluster sampling it is usually assumed that the clusters are disjoint. In this paper situations are considered in which this assumption is not fulfilled. Let the population π consist of N individuals "j", having the variates $V[j]$, $j = 1, 2, \cdots, N$, and let K clusters $C[i]$, $i = 1, 2, \cdots, K$, be such that each "j" belongs to at least one cluster. Let $s[j] \geq 1$ be the number of different clusters to which "j" belongs (the multiplicity of "j"). The cluster $C[i]$ contains N_i individuals with the variates $V[i, t]$, $t = 1, 2, \cdots, N_i$; $i = 1, 2, \cdots, K$. In a sampling procedure, let sub-sample sizes $n[i]$ be given for each $C[i]$, and weights $\lambda[i, t]$ for each $V[i, t]$; a random sample of k clusters $C[i_u]$, $u = 1, 2, \cdots, k$ is obtained, then $n[i_u]$ individuals are sampled from $C[i_u]$, and for each of them its variate and its multiplicity are recorded. Necessary and sufficient conditions are derived for $S = \Sigma_{u=1}^{k} \Sigma_{v=1}^{n[i_u]} V[i_u, t_v] \lambda[i_u, t_v]$ being an unbiased estimate of $\bar{V} = \dfrac{1}{N} \Sigma_{j=1}^{N} V_j$. The variance of S is found, the weights are studied which minimize this variance, and some practically important special cases are derived.

2. A Simple Nonparametric Test of Independence. Nils Blomqvist, University of Stockholm.

Consider a sample of size n from a two-dimensional distribution $F(x, y)$. Let $\tilde{x}$ and $\tilde{y}$ denote the two sample medians and compute the number of individuals, say k, satisfying the inequality $x < \tilde{x}$, $y < \tilde{y}$ (the trivial difficulty arising when n is an odd number can easily be overcome). A test of independence based on k is nonparametric. As a matter of fact one has under the null hypothesis that

$$P(k) = \binom{m}{k}^2 \Big/ \binom{2m}{m},$$

where $m = [n/2]$. In the case of normal F with correlation coefficient ρ it is possible to show, by studying the asymptotic behavior of the power function of the test in the neighborhood of $\rho = 0$, that the asymptotic efficienty of the test is $(2/\pi)^2$, or about 41% This result is based on the fact that k has an asymptotically normal distribution if some regularity conditions are fulfilled In spite of its low efficiency it is suggested that the test be used in cases where some information can be neglected in favor of the simplicity of the method.

3. On Minimax Statistical Decision Procedures and Their Admissibility. Colin R. Blyth, University of California, Berkeley.

The problem considered is that of using a sequence of observations on a random variable X to make a decision. Two loss functions W_1 and W_2, each depending on the distribution F of X, the number n of observations taken, and the decision δ made, are assumed given Minimax problems can be stated for weighted sums of W_1 and W_2, or for either one subject to an upper bound on the expectation of the other Under suitable conditions it is shown that solutions of the first type of problem provide solutions for all problems of the latter types, and that admissibility for a problem of the first type implies admissibility for problems of the latter types. Two examples are given: estimation of $E\,X$ when X is (1) normal with known variance, (2) rectangular with known range The two loss functions are in each case $W_1 = n$ and an arbitrary nondecreasing function $W_2(\,|\,\delta - \theta\,|\,)$ Admissible minimax estimates are obtained. Extensions to any function $W_1(n)$ are indicated, two examples are given for the normal case where the sample size must be randomised among more than a consecutive pair of integers.

4. Sufficient Statistics and Unbiased Estimates for "Selected" Distributions. Douglas G. Chapman, University of Washington, Seattle.

A family of distributions obtained from any given family by fixed selection may be called a "selected" family. Tukey's theorem that such selected families admit the same set of sufficient statistics as the parent family is proved for an extended class of distributions. Further if the selection does not involve truncation the existence of minimum variance unbiased estimates of parameters of the parent family ensures the existence of similar estimates for the selected family. Some results are derived for minimum variance unbiased estimates for truncated distributions.

5. The Unattainability of Certain Lower Bounds by Product Densities. R. C. Davis, U. S. Naval Ordnance Testing Station, China Lake.

Under weak regularity conditions it is shown that for the case in which the sample size is a nonrandom variable, certain lower bounds are unattainable. Consider a univariate chance variable X, possessing an absolutely continuous distribution function $F(x, \theta)$, in which θ is the unknown parameter Under quite general regularity conditions Barankin has proved the existence and uniqueness of the locally best unbiased estimate of a function $g(\theta)$ for a specified parameter value θ_0 The criterion of bestness is the minimization of the s^{th} absolute central moment ($s > 1$) of the estimate about $g(\theta_0)$, and Barankin has obtained an expression for the lower bound both in the general case and in particular for a case which yields a generalization of the Cramer-Rao inequality valid for any s^{th} absolute central moment It is the latter lower bound with which we are concerned With an additional weak assumption concer[illegible] density function of X, it is shown that if $\varphi_s(x_1, x_2, \cdots, x_n)$ is the locally best [illegible] θ_0) (obtained by Barankin) for each fixed sample size n and for each $s > 1$ [illegible] no probability distribution $F(x, \theta)$ except for $s = 2$ yielding a sequence $\{\varphi_s(x_1, x_2, \cdots, x_n)\}$ ($n = 1, 2, \cdots$, ad inf.) in which $x_1, x_2, \cdots, x_n$ are for each n independently and identically distributed chance

variables and for which $\varphi_s(x_1, x_2, \cdots, x_n)$ attains for each n the special lower bound given by Barankin. Obviously in the case $s = 2$, the lower bound is achieved by an efficient statistic if one exists.

6. A Note on the Power of the Sign Test. T. A. Jeeves and Robert Richards, University of California, Berkeley.

Values obtained by using the normal approximation to the noncentral t-distribution given by Johnson and Welch were compared with exact values given by Neyman and Tokarska. The comparison indicated that efficiencies of the sign test computed from the approximation would be consistently higher than the true efficiencies. To avoid this bias the sign test was randomized so that levels of significance of $\alpha = .05$ and $\alpha = .01$ were obtained and the exact values of the noncentral t used. Efficiencies were computed using various measures of equivalence of the power functions: (1) balancing the area (Walsh), (2) minimizing the maximum difference, (3) equalizing the power at certain fixed points. The various measures of equivalence yielded no marked differences in efficiencies. Tables were given of the efficiencies for small n. The efficiency for $\alpha = .05$ was about .7 for n between 6 and 20 and somewhat higher for $\alpha = .01$. The efficiency slowly approaches the asymptotic value of $2/\pi = .6366$ as n increases.

7. About Some Classes of Sequential Procedures for Obtaining Confidence Intervals of Given Length. (Preliminary Report). Werner R. Leimbacher, University of California, Berkeley.

The special class C_1 of such procedures indicated by A. Wald (*Sequential Analysis*, John Wiley & Sons, 1947, pp. 145–156) can be extended by generalizing and improving the inequality on which the procedures are based. It is shown that even in this larger class C_2, a procedure could possibly be optimum only under very special circumstances. The well known optimum procedure for a normal distribution $N(\theta, 1)$ can be obtained as the limit of a sequence of procedures from C_2. For the suggested sequence, however, the limit no longer belongs to C_2. In order to eliminate various deficiences of C_2, a modified class C_3 is proposed which contains the well known optimum procedures for the normal and rectangular distributions. The method indicated seems suggestive for the general case of estimating location parameters by confidence intervals.

8. On the Stochastic Independence of Symmetric and Homogeneous Linear and Quadratic Statistics. Eugene Lukacs, U. S. Naval Ordnance Testing Station, China Lake.

It is known that the sampling distributions of the mean and of the variance are stochastically independent if and only if the parent distribution is normal. This was proven by R. C. Geary (*Jour. Roy. Stat. Soc., Suppl.*, Vol 3 (1936)) and using a different method by E. Lukacs (*Annals of Math. Stat.*, Vol. 13 (1942)). The question arises whether there are any distributions having the property that the sampling distributions of the mean and of a symmetric and homogeneous quadratic statistic are independent. It can be shown that there are only the following possibilities: (1) the parent distribution is normal, (2) the parent distribution is degenerate with a single saltus of one, (3) the parent distribution is a step function with two steps, located symmetrically with respect to zero, (4) the parent distribution is a gamma distribution

9. The Distribution of the Maximum Deviation between Two Sample Cumulative Step Functions. Frank J Massey, Jr., University of Oregon.

Let $x_1 < x_2 < \cdots < x_n$ and $y_1 < y_2 < \cdots < y_m$ be the ordered results of two random samples from populations having continuous cumulative distribution functions $F(x)$ and

$G(x)$ respectively. Let $S_n(x) = k/n$ when k is the number of observed values of X which are less than or equal to x, and similarly let $S'_m(y) = j/m$ where j is the number of observed values of Y which are less than or equal to y. The statistic $d = \max_x |S_n(x) - S'_m(x)|$ can be used to test the hypothesis $F(x) \equiv G(x)$, where the hypothesis would be rejected if the observed d is significantly large. In this paper a method of obtaining the exact distribution of d for small samples is described, and a short table for equal size samples is included The general technique is that used by the author for the single sample case There is a lower bound to the power of the test against any specified alternative. This lower bound approaches one as n and m approach infinity proving that the test is consistent

10. An Iterative Construction of the Optimum Sequential Decision Procedure with Linear Cost Function. LINCOLN E. MOSES, Stanford University.

Where the cost of taking n observations is proportional to n, define a sequential decision procedure D_T by means of its associated "stopping region" T, T is the set of a posteriori probability distributions $\xi(\theta)$ for which D_T instructs the statistician to take no observation and to make the decision which minimizes the Bayes risk. Now let D_T be any sequential decision procedure which has uniformly bounded average risk for every a priori distribution, $\xi(\theta)$. Define T as the derived region of T T' is the set of $\xi(\theta)$ such that the Bayes risk of stopping at $\xi(\theta)$ is not greater than the risk of taking one observation and then using D_T. Define $T^{(n+1)} = T^{(n)'}$. Then it is shown that the sequence of regions $\{T^{(n)}\}$ $n = 1, 2, \cdots$ is monotonically decreasing to a limit region T^∞, and that D_{T^∞} is the optimum sequential decision procedure. Some numerical examples are given where the exact solution is obtained and the convergence of the iteration is examined. (This paper was prepared under the sponsorship of the Office of Naval Research.)

11. On the Law of the Iterated Logarithm for Dependent Random Variables. STANLEY W. NASH, University of California, Berkeley.

The order of the remainder term is evaluated in the distribution function of the asymptotically normal sum S_n of dependent random variables of a certain class considered by Loève. Bounds are found for the probability that $\max_{k \leq n} |S_n| \geq B_n^{\frac{1}{2}}x$, where B_n is the sum of the variances of components of S_n. Given an infinite sequence of events A_n, a necessary and sufficient condition is found for the probability that infinitely many A_n occur to equal one. This criterion extends criteria due to Borel. With these results established, the law of the iterated logarithm is shown to hold for a wide subclass of Loève's class of dependent random variables. Within this class the partial sum $S_n - S_i$ may approach normality with a speed which depends in a certain functional way on the previous sum S_i, and which may be arbitrarily slow for some values of S_i. The conclusions generalize earlier results due to W. Doeblin and N. A. Sapogov.

12. Conditional Expectation and the Efficiency of Estimates. PAUL G. HOEL, University of California, Los Angeles.

A probability density function, $f(x; \theta)$, is considered for which the range of x does not depend on θ and for which there exists a sufficient statistic for θ It is shown that under certain regularity conditions, there exists a unique unbiased sufficient estimate of θ among those sufficient estimates which can be expressed as functions of a particular sufficient statistic. This result, together with results of other authors, is used to show that for the class of statistics satisfying the regularity conditions, the method of Blackwell for improving an unbiased estimate of θ does not yield an essentially better estimate than a well known estimate.

13. **Optimum Estimates for Location and Scale Parameters.** RAYMOND P. PETERSON, University of California and National Bureau of Standards, Los Angeles.

Let $h_i(W \mid E, \theta) = W(\theta_i^*, \theta)p(E \mid \theta)$, where $p(E \mid \theta)$ is the joint probability density function of the n (not necessarily independent) sample values $x_1, \cdots, x_n$ which may be represented as a point $E = (x_1, \cdots, x_n)$ in the n-dimensional Euclidean sample space M. The unknown parameters, $\theta_1, \cdots, \theta_s$, may be represented as a point $\theta = (\theta_1, \cdots, \theta_s)$ in the s-dimensional Euclidean parameter space Ω. $W(\theta_i^*, \theta)$ is a real-valued, nonnegative, measurable weight function, defined for all E in M and θ in Ω, which represents the relative seriousness of taking the estimate $\theta_i^*(E)$ as the value of θ_i for any particular sample point E. Let $G(\theta)$ be the unknown cumulative distribution function of θ. Then $\theta_i^*(E)$ is defined to be a best estimate of θ_i, provided that, if $\theta_i(E)$ is any other estimate of θ_i in the class under consideration, $\bar{I} - I^* \geq 0$, where

$$I = \int_\Omega \int_M h_i(W \mid E, \theta) dE \, dG(\theta).$$

Let

$$r_i(\theta) = \int_M h_i(W \mid E, \theta) \, dE, \qquad \varphi_i(E) = \int_\Omega h_i(W \mid E, \theta) \, d\theta.$$

A general theorem is proved to the effect that if $h_i(W \mid E, \theta)$ is measurable over the product space $M \times \Omega$ and if $r_i(\theta)$ and $\varphi_i(E)$ are uniformly convergent integrals, then a best estimate $\theta_i^*(E)$ of θ_i exists provided that $r_i(\theta)$ is constant and that $\theta_i^*(E)$ minimizes $\varphi_i(E)$ for all points E in M. General methods are obtained for constructing best estimates for location and scale parameters, separately or jointly, and for functions of location and scale parameters from several populations. As special cases, results are derived which are analogous to converses of Theorems 1 and 2 in Kallianpur's, "Minimax Estimates of Location and Scale Parameters", Abstract, (*Annals of Math. Stat.*, Vol. 21 (1950), pp. 310–311).

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest.

Personal Items

Professor William Feller of Cornell University has been appointed Eugene Higgins Professor of Mathematics at Princeton University.

Dr. Leonard Kent, formerly on the staff at the University of Chicago in the School of Business, is now with the firm of Alderson and Sessions, 1905 Walnut Street, Philadelphia 3, Pennsylvania.

Dr. G. B. Oakland has resigned an associate professorship of statistics at the University of Manitoba to accept the position as Head of Biometrics Unit, Division of Administration, Department of Agriculture, Ottawa.

Dr. Norman Rudy has accepted an appointment as Assistant Professor at Sacramento State College, Sacramento, California.

Professor G. R. Seth has returned to India to accept the position of Professor of Statistics and Deputy Statistical Advisor to the Indian Council of Agricultural Research, New Delhi.

Mr. Eric Weyl, textile engineering consultant, formerly of Manchester, New Hampshire, has moved his office to 2509 Vail Avenue, Charlotte, North Carolina. Mr. Weyl, a specialist in cotton spinning, serves as regular consultant to many leading textile mills.

The completion and successful operation of SEAC—the National Bureau of Standards Eastern Automatic Computer—has been achieved by electronic scientists of the National Bureau of Standards. SEAC is a high-speed, general-purpose, automatically-sequenced electronic computer. It was developed and constructed, in a period of 20 months, by the staff of the National Bureau of Standards under the sponsorship of the Department of the Air Force to provide a high-speed computing service for Air Force Project SCOOP (Scientific Computation of Optimum Programs), a pioneering effort in the application of scientific principles to the large-scale problems of military management and administration. SEAC will also be available for solving important NBS problems of general scientific and engineering interest.

New Members

The following persons have been elected to membership in the Institute

(June 1, 1950 to August 31, 1950)

Aven, Russell E., M A. (Univ of Miss), Graduate student, University of Mississippi, *1511 North Main St., Water Valley, Mississippi.*

Bamberger, Gunter, Dip -Math (Univ. Gottingen), Division head in the Statistical Office of the City of Cologne, *Manderscheider Platz 12, Cologne-Sulz, Germany.*

Bangdiwala, Ishver S., M S (Univ N. C), Graduate student, University of North Carolina, *210 A Phillips Hall, University of North Carolina, Chapel Hill.*

Borch, Karl Henrik, M Sc (Oslo Univ), Field Science Officer for Middle East, *UNESCO, 19 Avenue Kieber, Paris 16e, France*

Buch, Kai R., M.Sc., Assistant Professor, Technical University of Denmark, *Eigaardsvej 14 A², Charlottenlund, Denmark.*

Carranza, Roque G., Ingeniero Industrial (Univ Buenos Aires), Consultant Industrial Engineer, *Parana 56, Buenos Aires, Argentina.*

Dominguez, Alberto G., Ph D. (Univ Buenos Aires), Professor of Mathematics, Facultad de Ciencias Exactas, Fisicas y Naturales, University of Buenos Aires, *Paraguay 1327, Buenos Aires, Argentina.*

Dunaway, William L., B S. (Univ. of Calif.), Graduate student, Dept. of Mathematical Statistics, University of California, *4320 Cahuenga Boulevard, North Hollywood, California.*

Fernandez, Jose J., Professor, University of Costa Rica, *Ap. 1313, San Jose, Costa Rica.*

Fortet, Robert, Ph.D (Paris), Professor, Department of Science de Caen, 168 Rue Caponiere, Caen (Caloados), France.

Geppert, Maria-Pia, Ph D (Univ of Giessen), Lecturer, University of Frankfurt; Head of Statistical Laboratory, Kerckhoff-Institute, Bad Nauheim; Lecturer, Technical High School, Darmstadt, Germany.

Gortler, J. Henry, Ph D. (Univ of Gottingen and Univ. of Giessen), Professor of Applied Mathematics and Dean of the Faculty of Natural Sciences and Mathematics, University

of Freiburg i. Br., Manager of "Gesellschaft fur angewandt Mathematik und Mechanik". *Stadtstrasse 57, Freiburg i Br, Germany.*

Guilbaud, George T., Agrege de l Univ (Paris), Chief, Section a l'Institute of Science Economique Appliquee, Paris, and Professor, Institute of Statistics, University of Paris, *35 Boulevard des Capucines, Paris 2, France*

Holloway, Clark, Jr., M.S. (Univ. of Ill.), Process Research Engineer, Gulf Research and Development Co, *P O 2038, Pittsburgh 30, Pennsylvania.*

Lieberman, Gilbert, M.A. (Columbia Univ.), Mathematician, U. S. Naval Research Laboratory, *220 Newcomb St., S.E, Washington 20, D. C.*

Lomax, K. S., M A. (Manchester Univ.), Lecturer in Economic Statistics, Economics Department, The University, Manchester, England.

Lorenz, Paul, Ph.D., Professor, University of Berlin, Kaiserstuhlstrasse 21, Berlin-Schlachtensee, Germany

Lunger, George F., M.B.A. (Univ. of Mich.), Statistician, Great Lakes Investigations, Fish and Wildlife Service, Department of the Interior, *2110 Arbor View Blvd., Ann Arbor, Michigan.*

Maggy, Robert K., M A. (Univ. of Calif.), Graduate student, University of California, *1685 Euclid Avenue, Berkeley 9, California.*

McElrath, Gayle W., M.S. (Univ. of Mich.), Assistant Professor, Department of Engineering, 208 Main Engineering Building, University of Minnesota, Minneapolis, Minnesota

Neisius, W. Vincent, M S. (Emory Univ.), Mathematics Instructor, Georgia Institute of Technology, *597 St. Charles Avenue, N.E., Atlanta 5, Georgia.*

Perloff, Robert, M.A. (Ohio State Univ.), Graduate student and Research Assistant, Research Foundation, Ohio State University, *1281 Bryden Road, Columbus 5, Ohio.*

Peter, Hans, Dr. rer. pol., Professor of Economics, University of Tübingen, *Tubingen-Wahlhausen 29, Germany.*

Putter, Joseph, M.Sc. (Hebrew Univ., Jerusalem), *International House, Berkeley 4, California.*

Rankin, Bayard, A.B. (Univ. of Calif.), Graduate student, University of California, *International House, Berkeley 4, California.*

Reid, Albert T., B.S. (Iowa State College), Research Assistant in Mathematical Biology, Committee on Mathematical Biology, University of Chicago, 5741 Drexel Avenue, Chicago 37, Illinois.

Shaw, Albert, B.S. (Univ. of Alberta), Lecturer, University of Alberta, Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada.

Shuhany, Elizabeth, A.M. (Boston Univ.), Assistant Instructor in Statistics and Assistant in Statistical Laboratory of Mathematics, Boston University, 725 Commonwealth Avenue, Boston 15, Massachusetts.

Stewart, John N., B.A. (Univ of Michigan), Graduate student, University of Michigan, *4834 Chalsworth, Detroit 24, Michigan.*

Strecker, Heinrich, Doctor der Naturwissenschaften (Univ. Munchen), Mathematical Statistician in the Bavarian Statistical Office, Rosenheimerstrasse 130, Munich 8, Germany.

Vaswani, Sundri (Miss) Ph.D. (Univ. of London), Research Associate in Statistics, c/o Ahmedabad Textile Industry's Research Association, P.O. Box 170, Ahmedabad, India.

REPORT OF THE BERKELEY MEETING OF THE INSTITUTE

The forty-fourth meeting of the Institute of Mathematical Statistics was held on August 5, 1950, on the Berkeley campus of the University of California, in conjunction with the Second Berkeley Symposium on Mathematical Statistics

and Probability which met from July 31 through August 12. Other organizations cooperating with the Symposium were the Biometrics Section of the American Statistical Association, The Western North American Region of the Biometric Society, the Econometric Society, the Institute of Transportation and Traffic Engineering of the University of California, and the Office of Naval Research. Some 218 persons registered for the Symposium, including the following 106 members of the Institute:

T. W. Anderson, Fred C. Andrews, Jane F. Andrian, Kenneth J Arrow, Edward W. Barankin, Helen P. Beard, Robert D Bedwell, Blair M Bennett, Joseph Berkson, Z. W Birnbaum, David Blackwell, E Blanco, Nils Blomqvist, Julius R Blum, Colin R. Blyth, A. H. Bowker, George W. Brown, Douglas G Chapman, C L Chiang, K L Chung, William G Cochran, Harald Cramér, Edwin L Crow, J. H Curtiss, R. C Davis, W J. Dixon, J L Doob, A. Dvoretzky, Mary Elveback, Benjamin Epstein, Mark W. Eudey, Edward A. Fay, William Feller, Edgar H Fickenscher, E Fix, William R Gaffey, Robert S Gardner, S. G Ghurye, M. A. Girshick, Paul Gutt, Jack C Gysbers, T E Harris, J L Hodges, Jr , Wassily Hoeffding, Paul G. Hoel, Harold Hotelling, John M Howell, Harry M. Hughes, R F. Jarrett, T. A. Jeeves, Mark Kac, Joseph Kampé de Fériet, E. S Keeping, Ryoichi Kikuchi, Wilfred M. Kincaid, H S Konijn, Charles H. Kraft, George M Kuznets, E. L Lehmann, Roy B Leipnik, Paul Levy, M. Loève, Arvid T. Lonseth, Eugene Lukacs, C. A Magwire, Jacob Marschak, Thomas Marschak, F. J Massey, Jr , A M Mood, Lincoln E Moses, James T. McWilliam, Stanley W Nash, J Neyman, Howard C. Nielson, Gottfried E Noether, Stefan Peters, John C. Petersen, Raymond P. Peterson, Robert I. Piper, Joseph Putter, Robert R. Putz, Bayard Rankin, Fred D. Rigby, David Rubinstein, Elizabeth L. Scott, Esther Seiden, Arthur Shapiro, Richard H Shaw, Ronald W. Shephard, W B Simpson, Monroe Sirken, M Sobel, Herbert Solomon, A. L Stewart, Donald E Stiling, G Szego, Robert Tate, William F Taylor, Leo J Tick, A W Tucker, Elizabeth Vaughan, Shanti A. Vora, Abraham Wald, Allen Wallis, J Wolfowitz, Miriam L Yevick.

Because of the extensive program of more than fifty invited addresses at the Symposium, the Institute meeting was devoted only to contributed papers. Professor David Blackwell of Howard and Stanford Universities presided at the Institute meeting, at which the following program was presented:

1. *Sampling from Populations with Overlapping Clusters* Z W Birnbaum, University of Washington, Seattle
2. *A Simple Nonparametric Test of Independence* Nils Blomqvist, University of Stockholm.
3. *On Minimax Statistical Decision Procedures and their Admissibility* Colin R Blyth, University of California, Berkeley
4. *Sufficient Statistics and Unbiased Estimates for "Selected" Distributions* Douglas G Chapman, University of Washington, Seattle
5. *The Unattainability of Certain Lower Bounds by Product Densities* R. C Davis, U S Naval Ordnance Testing Station, China Lake
6. *A Note on the Power of the Sign Test* T. A. Jeeves and Robert Richards, University of California, Berkeley
7. *About Some Classes of Sequential Procedures for Obtaining Confidence Intervals of Given Length* (Preliminary report) Werner R Leimbacher, University of California, Berkeley
8. *On the Stochastic Independence of Symmetric and Homogeneous Linear and Quadratic Statistics* Eugene Lukacs, U S Naval Ordnance Testing Station, China Lake

9. *The Distribution of the Maximum Deviation between Two Sample Cumulative Step Functions.* Frank J. Massey, Jr., University of Oregon.

10. *An Iterative Construction of the Optimum Sequential Decision Procedure with Linear Cost Function.* Lincoln E. Moses, Stanford University.

11. *On the Law of the Iterated Logarithm for Dependent Random Variables.* Stanley W. Nash, University of California, Berkeley.

12. *Conditional Expectation and the Efficiency of Estimates.* (By title). Paul G. Hoel, University of California, Los Angeles.

13. *Optimum Estimates for Location and Scale Parameters.* (By title). Raymond P. Peterson, University of California and National Bureau of Standards, Los Angeles.

The social activities at the Symposium included a tea on August 1, an excursion on August 3, a dinner on August 7, a picnic on August 9, and coffee on July 31 and August 2, 4, 7, 8, 10, and 11.

J. L. HODGES, JR.
Associate Secretary